

medikal istatistik

Doç. Dr. Oktay Özdemir



İstanbul
medikal
yayıncılık

Coefficients ^a		Standardized Coefficient <i>t</i> s	<i>t</i> Stat.	<i>p</i> Value
Unstandardized Coefficients	Std. Error			
.962	.930	.003	-1.056	.301
-.372	.930	-.372	-.387	.783



©İstanbul Medikal Yayıncılık BİLİMSEL ESERLER dizisi
Medikal İstatistik
Yazar: Doç. Dr. Oktay Özdemir
Omega Araştırma Şirketi İSTANBUL

1. Baskı 2005

ISBN - 975-6395-47-8

Yasalar uyarınca, bu yapıtin yayın hakları
İstanbul medikal yayıncılık ltd.şti.’ye aittir.
Yazılı izin alınmadan ve kaynak olarak gösterilmeden,
elektronik, mekanik ve diğer yöntemlerle
kismen veya tamamen kopya edilemez;
fotokopi, teksir, baskı ve diğer yollarla çoğaltılamaz.

2006 İstanbul Medikal Yayıncılık Ltd. Şti.
34104, Çapa-İstanbul-Türkiye

www.istanbultip.com.tr
e-mail: info@istanbultip.com

Deniz Abdal Mah., Başvekil Cad.,
Çeşme Sok., No. 3/Çapa, 34104 - İstanbul
Tel: 0212-584 20 60 - Faks: 0212 584 20 61

UYARI

Medikal bilgiler sürekli değişmekte ve yenilenmektedir. Standart güvenlik uygulamaları dikkate alınmalıdır, yeni araştırmalar ve klinik tecrübeler
içerdiği tedavilerde ve ilaç uygulamalarındaki değişiklıkların gereklili olabileceği bilinmelidir. Okuyucularla ilaçlar hakkında üretici firma tarafından
tanıttığından her ilaca alt en son ürün bilgilerini, dozaj ve uygulama şekillerini ve kontrendikasyonları kontrol etmeleri tavsiye edilir. Her hasta
ya **hasta** en iyi tedavi şeklini ve en doğru ilaçları ve dozlarını belirlemek uygulamayı yapan hekimin sorumluluğundadır. Yayıncı ve editörler bu
yayınlardan dolayı meydana gelebilecek hastaya ve ekipmanlara alt herhangi bir zarar veya hasardan sorumlu değildir.



yayına hazırlayan **istanbul medikal yayıncılık ltd. şti.**
imy adına sanat yöneticisi **özlem arabacı** 0212-584 20 60 / 13
yazar **oktay özdemir**
sayfa düzeni **aysegül oktay** 0212-584 20 60 / 13
qedaksiyon ve düzelti **murat kirti** 0216-339 00 55
kapak **imy tasarım**
baskı ve cilt **elma basım** 0212-670 05 25

Önsöz

İç Hastalıkları asistanlığımın ikinci yılında meslek hayatımın ilk yılını yapmak amacıyla kronik lenfositik lösemi hastalarının verilerini toplayıp, formlara kaydedip, yapılacak istatistik analizlerle ilgili destek almak üzere fakültenin Biyoistatistik Bölümü'ne başvurdum. Fakültede öğrenciyken gördüğüm Biyoistatistik derslerinin üzerinden yıllar geçmiş ve benim istatistikle ilgili hatırladığım şeyler çok azdı. Araştırma ile ilgili yapılacak istatistik analizlerle ilgili konulara geçmeden önce, uzunca bir süre konu ile ilgili medikal terimleri anlatmaya çabalamış ve sonunda da birkaç kere daha giđip gelerek, nihayet araştırmamın istatistik analizlerini yapabilmiştim.

İkinci araştırmamın analizi sırasında da benzeri iletişim sorunları yaşamınca, bir doktorun kendine yetecek kadar istatistik bilmesi gerektiğini düşünüp, kendimi geliştirmeye karar verdim. Önce biraz kitap karıştırdım, sonra birkaç kursa gittim, daha sonra bir yan dan asistanlığa devam ederken, öte yandan özel öğrenci statüsünde Biyoistatistik ve Epidemiyoloji konularında master programlarına devam ettim. Sonunda geçen yıllarla birlikte, matematiği seven bir doktor olmanın avantajı ile temel biyoistatistik konularını öğrendim ve medikal istatistik konusunda ders ve kurslar verecek kadar kendimi geliştirdim.

Bu birkimimi tipta istatistik kullanımını konularını içeren bir kitapla, tüm sağlık camiası ile paylaştım. Doktorların istatistik analizle ilgili yaşadıkları sorunların çözümü ile ilgili ihtiyaçlarının farkında olarak, 1995 yılında tipta bilgisayar kullanımı konusunda çok deneyimli olan, o zamanki arkadaşım ve şimdi hem arkadaşım, hem de iş ortağım olan Doç. Dr. Murat Hayran'la birlikte, istatistikle ilgili bölümlerini benim yazdığım "Bilgisayar, İstatistik ve Tıp" kitabını yayınladık.

Doktorların yayım yapmaya gayret ederken en büyük sıkıntısı, araştırmalarının istatistik analizlerinin planlanması ve yapılmasıdır. Önceleri kağıt kalemlle, daha sonraki yıllarda elektronik hesap makineleri yardımıyla istatistik analizleri yapmaya çalışırken, artık çok ayrıntılı istatistik analiz yazılımları sayesinde, hem çok daha hızlı, hem de çok daha fazla seçenekleri kullanarak, sorunlarımızı çözmeye çalışıyoruz. Ama yine de sorunlar bitmiyor. Çünkü, bilgisayarlar çok hızlı, ama (en azından şimdilik) hangi durumlarda hangi istatistik testlerin kullanılacağı konusunda çok yardımcı olmuyorlar. Bu yüzden elimizdeki çeşitli bilgisayar programlarına rağmen, öncelikle doğru istatistik test seçimini kendimizin yapması gerekiyor.

Bu kitabın amacı, teorik istatistik bilgileri aktarmak değil, medikal istatistik ile ilgili temel kavramları uygulamalı olarak anlatmaya çalışmaktadır. Böylece üzerinde çalışığınız araştırmaya en uygun analiz seçenekleri konusunda doğru karar verebilecek, okuduğunuuz makalelerdeki sonuçları doğru yorumlayabileceksiniz.

Bu kitabın hazırlanmasında bana en önemli katkıyı, istatistik analiz konusunda benden destek isteyen hekim arkadaşlar, dernekler ve kurumlar sağladı. Analizinde yardımcı olduğum her araştırmadan çok şey öğrendim, öğrenmeye devam ediyorum.

Çok sayıda şekil, grafik, tablo ve rakam içeren bu kitabın içeriğinin düzenlenmesinde tüm çabaşkanlığıyla bana yardımcı olan Murat Kirtiş'in katkıları da vurgulanmaya değer, Murat'a teşekkür ederim...

Yazılması iki yıla yakın süre alan bu kitabı yayınlanması için beni cesaretlendiren, İstanbul Medikal Yayıncılık'ın yöneticiSİ İsmail Şahin bey'e ve titiz, dikkatli ve özverili çalışmaları ile kitabı yama hazırlayan ekibinden Özlem Arabacı ve Ayşegül Oktay'a teşekkürü borç bilirim...

En büyük teşekkür ise eşime... Aytül'ün, yalnızca bu kitabı yazarken uykusuz geçirdiğim gecelerde bana gösterdiği sabır ve hoşgörüsü ve sırtımı sivazlaması değil, hayatımın her anında bana hissettirdiği sevgisi ve sınırsız desteği olmasaydı, bu kitabı hayatı geçirememedim. En içten teşekkürlerim eşime...

Kitabın hazırlanmasında gösterilen tüm titizliğe rağmen, gözden kaçmış olabilecek hatalar Görürseniz lütfen iletin. Kitapla ilgili bildireceğiniz görüş ve eleştirileriniz, kitabın sonraki baskıları için çok yararlı olacaktır.

Umarım bu kitaptan yararlanırsınız.

Oktay Özdemir
Eylül 2005

Özgeçmiş

Oktay Özdemir, 1962'de Ankara'da doğdu. ÖSS Türkiye birincisi olarak girdiği Hacettepe Üniversitesi Tıp Fakültesi'den 1986'da mezun oldu. 1991'de İç Hastalıkları, 1994'de Hematoloji uzmanı oldu. 1995'de Hematoloji Doçenti unvanını aldı. 1996'da Hacettepe Üniversitesi Klinik Araştırmalar Ünitesi'nin kurulması çalışmalarında yönetici olarak görev aldı. 1997'de üniversitedeki görevinden ayrılp, Omega Araştırma Şirketi'ni kurdu.

Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü'nde 1993-94 dönemi "Eksperimental Tasarımlarda İstatistik", 1994-95 dönemi "Temel Epidemiyoloji İlkeleri" ve 1994-95 dönemi "Epidemiyolojide Araştırma Tasarımları" master programlarına katıldı. 1995 yılında EORTC'nin (European Organization for Research and Treatment of Cancer) "Biyoistatistik", 1996 yılında PERI'nin (Pharmaceutical Education and Research Institute) "Farmakoekonomi" eğitim programlarına katıldı.

Çeşitli tıp dergilerinde editör, yardımcı editör, editörler kurulu üyesi olarak görev'e aldı. Uluslararası hakemli dergilerde 40 ve ulusal dergilerde 50 olmak üzere 90 yayım, uluslararası kongrelerde 70 ve ulusal kongrelerde 60 olmak üzere 130 kongre bildirisini (sözlü/poster) vardır. Çeşitli kongre ve sempozyumlarda yaklaşık 20 oturumda panel katılımcısı ya da konferansçı olarak görev almıştır.

Kurucu ortağı ve yöneticiSİ olduğu Omega Araştırma Şirketi'nin dokuz yıla yakın deneyimi içinde yaklaşık 450 araştırma projesinde çalışmıştır.

Halen Omega'nm kurucu ortağı ve Veri Yönetimi ve İstatistik Departmanı yöneticisidir. Omega'nm hazırladığı "Medikal İstatistik", "Araştırma Teknikleri", "Örneklem Büyüklüğü Hesaplanması", "Kanita Dayalı Tıp" ve benzeri konulardaki eğitim programlarının koordinatörü ve eğitmcisi olarak, hekimler, dernekler ve ilaç firması çalışanlarına yönelik 30 civarında eğitim programında 1000'e yakın hekimin yüzüze eğitiminde yer almıştır.

İçindekiler

■ Tıpta çalışma düzenleri	1
■ Örneklemme ve randomizasyon	15
■ Değişkenler ve özellikler	33
■ Veritabanının hazırlanması	47
■ Tanımlayıcı istatistik	61
■ Analitik istatistik	83
■ Sayısal veriler için tanımlayıcı grafik ve tablolar	95
■ Ortalamaların karşılaştırılması	107
■ Tek grup ortalaması	115
■ İki ortalamanın karşılaştırılması	123
■ Ikiden fazla ortalamanın karşılaştırılması	141
■ Oranların karşılaştırılması	177
■ Bağıntı analizi	213
■ Regresyon analizi	227
■ Sağkalım analizi	249
■ Çok değişkenli istatistik analiz yöntemleri	277
■ Tıpta karar verme	301

TİPTA ÇALIŞMA DÜZENLERİ

bölüm 1

Gözlemsel çalışmalar

Tanımlayıcı çalışmalar (vaka raporları)

Retrospektif çalışmalar

Kesitsel çalışmalar

Prospektif çalışmalar

Deneysel çalışmalar

Bağımsız eşzamanlı kontrollü çalışmalar

Kendi kendine kontrollü çalışmalar

Çapraz çalışmalar

Diş kontrollü çalışmalar

Ne zaman hangi çalışma düzeni seçilmeli?

Tıp alanındaki çalışmaların amacı, hastalıkların neden, sıklık, risk etkenleri, прогноз gibi nitelikler açısından tanımlanması, hastalık tanılarının konulması ve tedavilerinin yapılması için gerekli verilerin toplanmasıdır. Her amaca uygun çalışma düzenleri farklıdır. Bazen de eldeki olsanaklar nedeniyle amaca tam uygun olmayan bir çalışma düzeni seçilmek zorunda kalınabilir.

Çalışma düzenleri genel olarak gözlemsel ve deneysel olarak ikiye ayrılabilir. Gözlemsel çalışmalar genellikle hastalıkla ilgili neden, sıklık, risk etkenleri, прогноз gibi nitelikleri belirlemeye yarar. Deneysel (ya da klinik) çalışmalar ise tanrı ve tedavi ile ilgili yöntemlerin denendiği çalışmalarlardır.

GÖZLEMSEL ÇALIŞMALAR

Tanımlayıcı, retrospektif, kesitsel ve prospektif olmak üzere başlıca dört tip gözlemsel çalışma düzeni vardır.

En basit çalışma düzeni olan tanımlayıcı çalışmada genellikle küçük bir grup denekte (hasta, fare, akciğer filmi, böbrek biyopsisi, vb.) göze çarpan, ilginç bulgular rapor edilir. Bir çok hastalığın tanımlanmasında (Behçet hasta-

lığı gibi), tanısında ve tedavisinde fitili ateşleyen bu tip çalışmalarlardır.

Diğer üç tip gözlemsel çalışma düzenini birbirinden ayıran, çalışmanın süresi ve deneklerde araştırılan konunun çalışmanın başlangıç zamanından önce, aynı anda ya da sonra olduğunu söyler. Prevalans çalışmaları adı da verilen kesitsel çalışmalarında deneklerin çalışma anındaki verileri değerlendirilir. Prospektif çalışmalarında çalışma yönü ileriye dönüktür; örneğin bir risk etkeninin varlığında bir hastalığın ortaya çıkması izlenir. Retrospektif çalışmalar ise çalışmanın yönü geriye dönüktür; yani bir hastalık durumunda geriye dönerken, hastalığa neden olabilecek risk etkenleri aranır.

Tanımlayıcı çalışmalar (vaka raporları)

Genellikle bu tip çalışmalar küçük bir grup hastada gözlenen ilginç sonuçların bildirilmesi şeklindekidır. Hastaların izlem süresi kısadır ve kontrol grubu yoktur. Bu çalışmalarla baştan planlanmış bir çalışma düzeni yoktur ve genellikle istatistik analiz de yer almaz. Yine de değerlensiz çalışmalardır denemez. Çünkü bu çalışmalarla bildirilen ilginç bulgular, daha sonra prospektif çalışmaların planlanması刺激する。

Benzenle temas eden beş kişide lösemi ortaya çıktığının bildirilmesi tanımlayıcı çalışmaya bir örnektir.

Retrospektif çalışmalar

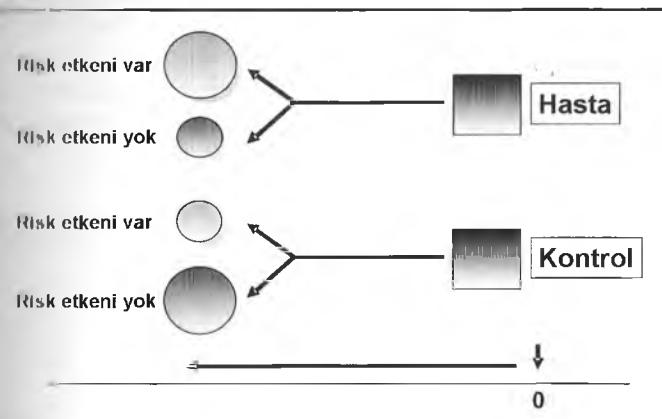
Retrospektif düzende (*case-control study*) araştırılan soru "Ne oldu?" sorusudur. Bu çalışmalarla hasta olan (*case*) ve olmayan (*control*) gruplar ele alınarak, bu deneklerin geçmişlerinde hastalığa neden olabilecek risk etkenleri araştırılır. Retrospektif çalışma, genellikle anlaşıldığı gibi hastaların dosyalarının ya da poliklinik kayıtlarının değerlendirildiği çalışmalar değildir.

Örnek 1'deki çalışma düzeni, retrospektif çalışma düzenidir (Şekil 1).

ÖRNEK 1

Lösemiye neden olan risk etkenleri araştırılmak isteniyor ve lösemisi olan 100 kişi (*case*) ve ayrıca 100 sağlıklı kişiye (*control*) özgeçmişlerinde olası risk etkenlerinin var olup olmadığı soruluyor. Hastaların 40'ında ve sağlıklı kontrollerin 10'unda özgeçmişinde benzenle temas olduğu saptanıyor. Bu bulgu benzinin lösemi için bir risk etkeni olduğunu düşündürür.

Uygun istatistik analiz yöntemleri ile bu farkın gerçekten önemli olup olmadığı test edilerek, lösemi risk etkenleri aydınlatılmış olur.



Şekil 1
Retrospektif çalışma düzeni.
Aşağıdaki kalın ok çalışmanın yönünü, ince ok zamanı ve "0"
çalışmanın başladığı anı
temsil etmektedir.

Retrospektif çalışma düzeninin en önemli zorluğu ya da sakıncası geçmişteki olayların tanımlanması ile ilgili güçlüklerdir. Örneğin "benzenle temas" olup olmadığıının kesin olarak belirlenmesi kolay olmayabilir. Benzenle teması olmadığını söyleyen bir kişinin kendisinin de farketmediği ya da önem vermediği için sonradan unuttuğu bir temas öyküsü olabilir.

Kesitsel çalışmalar

Kesitsel çalışma düzende (*cross-sectional study*) araştırılan soru "Ne oluyor?" sorusudur. Bu çalışma düzende denekler ya hiç izlenmezler ya da çok kısa bir süre izlenirler. Genellikle hastalık sıklığı (prevalansı), hastalık mekanizmaları ve tanışal sorunları çözümeyi amaçlayan çalışmalarlardır. İzlem süresi kısa olduğu için çabuk sonuçlanırlar.

ÖRNEK 2

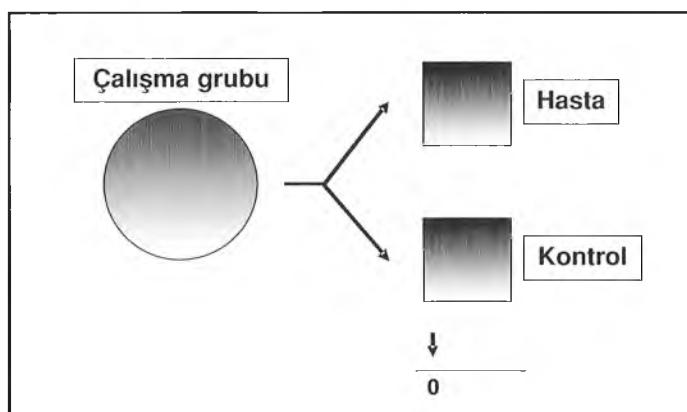
Türkiye'de kalp hastalığına neden olan hipertansiyon, hiperlipidemi, diabetes mellitus gibi hastalıkların prevalansı, yaş ve cinsiyetle ilişkisinin araştırıldığı TEKHARF çalışması, kesitsel prevalans çalışmalarına bir örnek olarak verilebilir. Bu çalışmada uygun örneklemeye yöntemleri ile seçilen kişilerde, öykü, fizik muayene ve laboratuvar çalışmaları ile hipertansiyon, hiperlipidemi ve diabetes mellitus prevalansı araştırılmıştır.

Örnek 2'deki TEKHARF çalışmasının ilk aşamasında hastalar bir kez görülmüş ve izlenmemiştir. İzlem süresinin çok kısa olması ya da olmaması, kesitsel çalışmaların özellikleidir.

Kesitsel çalışmalar, hastalık mekanizmaları ve tanısal sorunların çözümü amacıyla da yapılır. Örnek 3'teki çalışma, bu tür bir kesitsel çalışmaya örmektir (Şekil 2).

ÖRNEK 3

Akut renal transplant rejeksyonu ile siklosporin nefrotoksitesini birebirinden ayırmak için idrar β_2 -mikroglobulin düzeyinin yararlı olup olmadığı araştırılmak isteniyor. Böbrek fonksiyonları bozulan transplant hastalarından idrar alınıp β_2 -mikroglobulin düzeyi bakıldıktan sonraki günler içinde bu iki klinik durum, klinik gidiş ya da greft biyopsisi gibi başka bir parametre ile birebirinden ayrıldığında, hastanın izlem süresi bitiyor. İki gruptaki idrar β_2 -mikroglobulin düzeyleri uygun istatistik yöntemle karşılaştırıldığında, iki klinik durumun ayrıntı tanısında idrar β_2 -mikroglobulin düzeyinin yararlı olup olmadığı anlaşılır.



Şekil 2

Kesitsel çalışma düzeni. Aşağıdaki ince ok zamanı ve "0" çalışmanın başladığı ami temsil etmektedir.

Prospektif çalışmalar

Prospektif çalışma döneminde (kohort çalışma: *cohort study*) araştırılan soru "Ne olacak?" sorusudur. Bu çalışma döneminde retrospektif çalışma döneminin tersine, çalışmaya alınan tüm denekler başlangıçta sağlıklıdır. Bu denekler, araştırmanın konusu olan hastalığın olası risk etkenlerine sahip olup olmamalarına göre iki grup halinde belli bir süre izlenirler. Sürenin sonunda

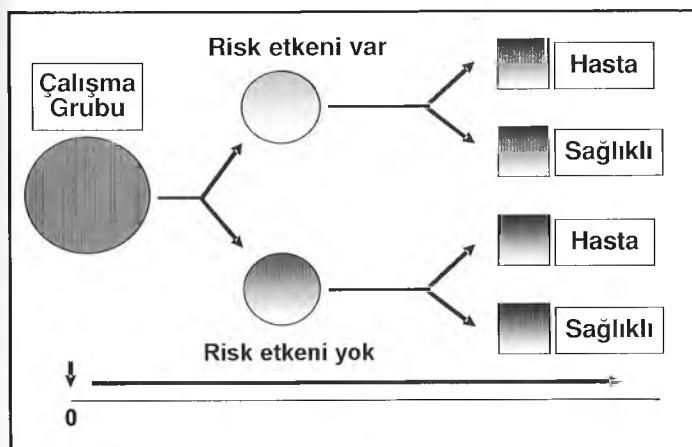
risk etkenine sahip olan ve olmayan deneklerde hastalığın ortaya çıkma sıklıkları (insidans) karşılaştırılarak, olası risk etkenin gerçekten de önemli olup olmadığı anlaşılır.

Örneğin daha önceki vaka raporları ve retrospektif çalışmalardan, benzen ile lösemi arasında ilişki olduğu sonucunun çıktığini, ama bu çalışma düzenleri ile ilgili sözü edilen sakincalar nedeniyle konunun açıklık kazanmadığını düşünelim. Bunu daha kesin anlamanın yolu prospektif düzende bir çalışma yapmaktadır.

Örnek 4'deki çalışma düzeni, prospektif kohort çalışma düzeninin tipik bir örneğidir (Şekil 3).

ÖRNEK 4

Löseminin etyolojisinde benzenin rolünü ortaya koymak amacıyla sağlıklı (lösemisi olmayan) 200 kişi, 20 yıl süreyle izleme alınıyor. Yirmi yıl boyunca bu deneklerden 100'ünde benzenle temas olmadığını ve diğer 100 kişinin ise meslekleri gereği benzenle teması olduğunu varsayıyalım. 20 yıl sonra benzenle teması olmayanların ikisisinde, teması olanların sekizinde lösemi gelişir ve uygun istatistik yöntemlerle aradaki farkın önemli olduğu anlaşılırsa, benzen-lösemi ilişkisi kesinleşmiş olur.



Şekil 3
Prospektif kohort çalışma düzeni.
Aşağıdaki kalmak ok çalışmanın yönünü, ince ok zamanı ve "0" çalışmanın başladığı anı temsil etmektedir.

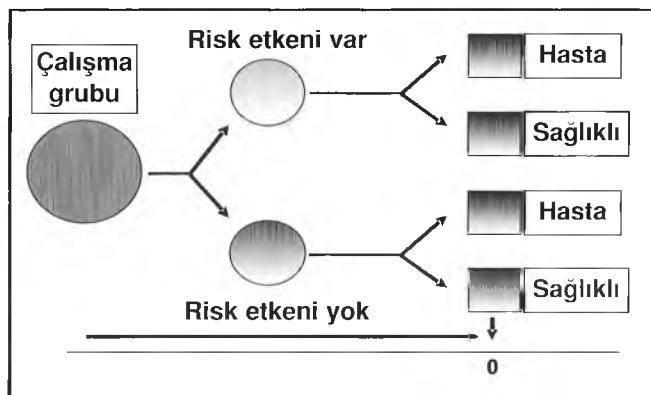
Prospektif çalışmaların araştırılan soruya net yanıt bulmak gibi önemli bir avantajı yanında, düzenli denek izlemi ve uzun izlem süresi nedeniyle geç sonuçlanması gibi sakincaları vardır.

Retrospektif kohort çalışmalar

Retrospektif çalışmalar hasta dosyalarının incelenmesinden ibaret olmadığı gibi, her dosya taraması şeklinde yapılan çalışma da retrospektif değildir.

Kohort çalışmalarında, genellikle çalışmanın başlangıç zamanı şimdiki zaman, bitiş zamanı gelecek zamanken, bazan başlangıç zamanı geçmiş zaman, bitiş zamanı şimdiki zaman olabilir. Bu tip çalışmalarla retrospektif (*historical*) kohort çalışmalar adı verilir (Şekil 4). Retrospektif kohort çalışmalar, prospektif kohort çalışmalarla karşılaşıldığında, kısa sürede sonuçlanması nedeniyle daha avantajlıdır.

Şekil 4
Retrospektif kohort çalışma düzeni.
Aşağıdaki kalın ok çalışmanın yönünü, ince ok zamanı ve "0"
çalışmanın başladığı anı temsil etmektedir.



Örnek 5'de izlem süresinin başlangıcında tüm deneklerin sağlıklı olması ve risk etkeni (sigara) olup olmamasına göre gruplandırılmasının nedeniyle, bu çalışma kohort tipi bir çalışmadır. Ancak prospektif kohort çalışmalarında, denek izleniminin başlangıcı şimdiki zamanda, izlem süresinin bitimi gelecek zamandadır. Oysa bu çalışmada 20 yıllık izlem süresi geçmiş zamanda, izlem süresinin bitimi ise şimdiki zamandadır. Bu nedenle bu tip kohort çalışmaları, retrospektif kohort çalışmalar olarak adlandırılır.

Retrospektif kohort çalışması yapabilmek için hasta kayıtlarının tam olması gereklidir. Bu nedenle sağlık kayıtlarının çok yetersiz ve düzensiz olduğu Türkiye'de çok nadiren yapılabilecek tipte bir çalışma düzenidir.

Hem retrospektif (*case-control*) çalışmaların, hem de prospektif kohort çalışmaların amacı hastalık risk etkenleri ve nedenlerinin ortaya konulmasıdır. Retrospektif çalışmalarında hastalıklı grupta geriye dönük olarak risk etkenleri

ÖRNEK 5

Son 20 yıla ait kayıtların eksiksiz olduğu bir hastanede, bu kayıtlardan yararlanarak, 20 yıl önce sağlıklı olduğu bilinen kişilerden, 20 yıl boyunca sigara içmemiş olan 500 ve içmiş olan 500 olmak üzere toplam 1000 kişinin adreslerine mektup yazılarak çağrılıyor. Gelenlere koroner anjiyografi yapılarak koroner arter hastalığı açısından değerlendiriliyor. Bu süre içinde ölmüş olan kişilerin ölüm nedenleri arasında koroner arter hastalığı olup olmadığı da hastane kayıtlarından ya da mektuplara gelen yanılardan anlaşılıyor. Sigara içenlerde ve içmeyenlerde 20 yıl içinde koroner arter hastalığı ortaya çıkma sıklığı karşılaştırılıyor.

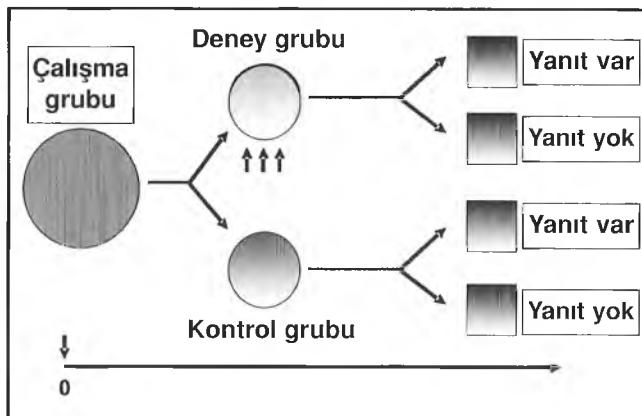
arastırılırken, prospektif çalışmalarında risk etkeni taşıyan sağlıklı bireylerde belirli bir sürede hastalığın ortaya çıkma sıklığı araştırılır. İyi düzenlenmiş prospektif çalışmalarla, retrospektif çalışmalarla göre daha doğru bilgi edinilir. Çünkü retrospektif çalışmalarla ilgilenilen risk etkenleri dışında da gözden kaçan ya da kontrol edilemeyen etkenler olabilir. Öte yandan retrospektif çalışma düzeni kısa zamanda sonuçlanması ve daha ucuz mal olması gibi nedenlerle sıklıkla kullanılmaktadır. Ayrıca nadir görülen hastalıklarda prospektif çalışmaların süresi, hastalığın ortaya çıkması için yeterli olmayabilir.

DENEYSEL ÇALIŞMALAR

Deneysel çalışmalar genellikle tanı ve tedaviyle ilgili yöntemlerin etkinliklerinin denendiği çalışmalarlardır. Kontrollü ve kontrollsız çalışmalar olmak üzere ikiye ayrılır. Kontrollü çalışmalarla etkinliği araştırılan yöntem, placebo ya da başka bir yöntemle karşılaştırılır. Kontrollsız çalışmalar ise hastaların hepsine aynı yöntem uygulanır ve sonuçları belirtilir. Bilimsel çalışmalarla amaç bir girişimin (tedavi, tanı yöntemi vb.) bilinen yöntemlerden farklı olup olmadığını belirlenmesi olduğuna göre, kontrollsız çalışmalar pek değerli değildir.

Kontrollü çalışmalarla, etkinliği araştırılan yöntemin yerine placebo ya da başka bir yöntemin uygulanacağı grup (kontrol grubu) çeşitli şekillerde seçilebilir.

Şekil 5
Bağımsız kontrollü klinik çalışma düzeni.
Aşağıdaki Kahn ok çalışmamın yönünü ve "0" çalışmamın başladığı anı temsil etmektedir.



Bağımsız eşzamanlı kontrollü çalışmalar

Kontrollü çalışmalarında sıklıkla kullanılan bir yöntem, bağımsız eşzamanlı kontrol grubu oluşturmaktır. Bu düzende iki grup denk vardır. Deney grubuna (*experimental group*) araştırma konusu olan tanı ya da tedavi yöntemi uygulanırken, kontrol grubuna placebo ya da diğer yöntem uygulanır (Şekil 5).

Kontrol grubunun uygulanan yöntemdeki farklılık dışında deney grubuna benzer olması gereklidir. Zaten kontrol grubunun eşzamanlı olarak çalışmaya alınmasının amacı budur. Hastanın ya da araştıracının bir hastaya uygulanan yöntemin ne olduğunu bilmesi bazen sonuçların değerlendirilmesinde önyargiya yol açabilir. Eğer hem hastalar, hem de araştırcı hangi hastanın hangi gruba alındığını bilmiyorsa, bu çalışmalar çift-kör (*double-blind*) çalışmalar olarak tanımlanır. Yalnızca hastalar, kendilerine hangi yöntemin uygulandığını bilmiyorlarsa, tek-kör çalışma adı verilir. Hastaların iki gruptan hangisine ve nasıl alınacağı grupların benzerliğinin sağlanması açısından çok önemlidir. Hastaların gruplara alınma yöntemine göre bu çalışma düzeneğini ikiye ayırlır: randomize çalışmalar ve randomize olmayan çalışmalar.

Randomize çalışmalar, tüm araştırma düzenleri içinde en güvenilir sonuç elde edilen çalışmalarıdır. Bir çalışmada deney ve kontrol gruplarının belirlenmesi tesadüfi değilse, randomize olmayan çalışma adı verilir. Bu çalışma düzende hastaların gruplara seçilmesinde, grupların benzerliğini bozan etkenler olabilir.

Örnek 6'daki çalışma, randomize kontrollü çalışma içinde yapılmıştır.

ÖRNEK 6

Akut akciğer tromboembolisinin tedavisinde klasik tedavi olan heparinle yeni bir tedavi seçeneği olan ürokinazın etkinliğinin karşılaştırıldığı bir çalışmada, çalışmaya alınan her yeni hastaya ürokinaz mı, yoksa klasik tedavi mi verileceği yazı-tura ile belirleniyor. Daha sonra hastalara belirli aralarla perfüzyon-ventilasyon sintigrafisi ve pulmoner anjiografi yapılarak rekanalizasyon süresi saptanıyor ve iki tedavi yönteminin arasında fark olup olmadığı araştırılıyor.

ÖRNEK 7

Örnek 6'daki çalışmada etik nedenlerle yaygın tromboembolisi olanlarda henüz etkili olup olmayacağı bilinmeyen ürokinaz yerine heparin veriliyor. Daha hafif klinik durumlarda ise hastaya ürokinaz mı, yoksa klasik tedavi mi verileceği yazı-tura ile belirleniyor. Daha sonra hastalara belirli aralarla perfüzyon-ventilasyon sintigrafisi ve pulmoner anjiografi yapılarak rekanalizasyon süresi saptanıyor ve iki tedavi yönteminin arasında fark olup olmadığı araştırılıyor.

Örnek 7'deki çalışma düzeni, randomize olmayan çalışma düzenidir. Bu durumda heparin verilen grupta, alitta yatan klinik durumun ağırlığı nedeniyle ölüm oranı daha yüksek olabilir ve bu nedenle ürokinazın etkinliğini gerçekle olduğundan daha yüksek olarak saptanabilir. Bu nedenle, randomize olmayan çalışmaların sonuçları güvenilir olmayı bilir.

Etkili olup olmayacağı henüz belli olmayan yeni yöntemlerin denenmesinde grupların belirlenmesinde randomizasyon yapılması, çalışmanın sonuçlarının sağlıklı olması açısından çok önemlidir. Öte yandan Örnek 7'de olduğu gibi, ciddi klinik durumlarda yeni yöntemin uygulanması, özellikle kör çalışmalararda, etik sorunlar doğurabilir. Bu nedenle zaman zaman yapılan incelemelerde, yöntemlerden birinin diğerinden daha üstün olduğu kesinleşirse, çalışma planlanan süreden önce kesilerek, olası etik sorunlar önlenmiş olur.

Kendi kendine kontrollü çalışmalar

Deney ve kontrol gruplarında farklı hastalar bulunduğu zaman, randomizasyona rağmen, gruplar arasında istenilen benzerlik sağlanamayabilir ya da çalışmanın yapıldığı dönemde henüz bilinmeyen etkenler açısından randomizasyon yapılmadığı için, daha sonraki bulgular nedeniyle randomizasyon yönteminin geçersiz olduğu anlaşılabılır. Bu sorunun uygun bir çözümü aynı hastaların hem deney, hem de kontrol grubunda yer alması ile sağlanabilir. Bu tür çalışmalara kendi kendine kontrollü (*self-control*) çalışmalar adı verilir.

Örnek 8'deki çalışma düzeni, bağımsız kontrollü çalışma düzenidir. Örnek 9'da ise kendi kendine kontrollü çalışma düzeni seçilmiştir.

ÖRNEK 8

Diyetin serum kolesterol düzeyine etkisini araştırmak amacıyla yapılan bir çalışmada polikliniğe başvuran hastaların yarısına normal, yarısına yağıdan fakir diyet verildikten üç ay sonra serum kolesterol düzeyine bakılıyor.

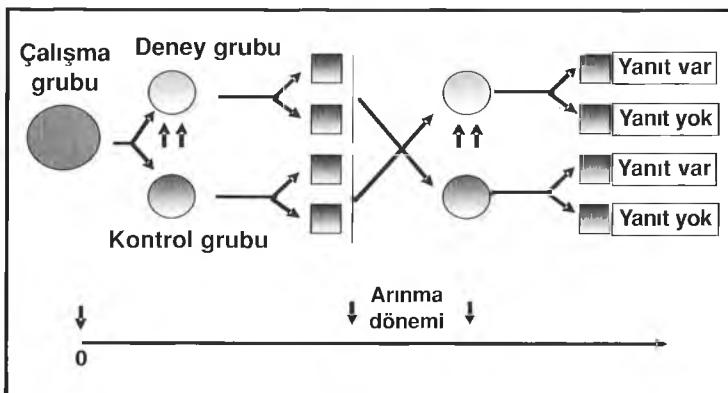
ÖRNEK 9

Diyetin serum kolesterol düzeyine etkisini araştırmak amacıyla yapılan bir çalışmada polikliniğe başvuran, normal diyetle beslenen hastalarda serum kolesterol düzeyine bakılıyor. Daha sonra yağıdan fakir diyete başlanıyor ve üç ay sonra serum kolesterol düzeyine tekrar bakılıyor.

Örnek 8'deki bağımsız kontrollü çalışma döneminde, şişmanlık, ailesel lipid metabolizması gibi bireysel farklılıklar iyi kontrol edilmezse, gruplardan biri diğerinden bu yönlerden farklı olabilir. Bu nedenle serum kolesterol düzeyine diyet dışındaki etkenlerin etkileri işi karıştırabilir. Kendi kendine kontrollü düzende ise, her iki tür diyetin verildiği gruplar aynı hastalardanoluştuğu için bu sorun ortaya çıkmaz. "Önce-sonra" tipindeki bu çalışma düzeni, oldukça sık kullanılmaktadır.

Çapraz çalışmaları

Çapraz (*crossover*) çalışma döneminde, hem bağımsız kontrol, hem de kendi kendine kontrol grupları vardır. Başlangıçta hastalar, bağımsız iki grup oluşturmak üzere ikiye ayrılır. Birinci gruba araştırma konusu olan yöntem, diğer gruba ise placebo ya da diğer yöntem uygulandıktan sonra, bir süre ara verilir. Bu araya, arınma dönemi (*washout period*) denir. Daha sonra birinci gruba placebo, ikinci gruba asıl yöntem uygulanır (Şekil 6).

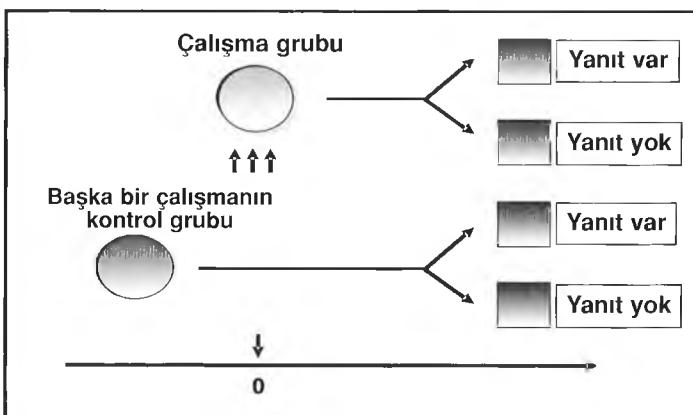


Şekil 6
Çapraz klinik çalışma düzeni.
Aşağıdaki kalın ok çalışmanın yönünü ve "0" çalışmanın başladığı anı temsil etmektedir.

Çapraz çalışmaları sonuçları oldukça güvenilir ve güçlü çalışmalarlardır. Ancak her klinik durumda uygun olmayabilir. Örneğin iki farklı ilaçın etkinliği karşılaştırılırken, eğer ilaçlardan birinin, uzun süreli etkisi varsa, ilk önce bu ilaçın verildiği grupta, aradaki arınma dönemi yeterli olmamıştır ve bu gruba diğer ilaç uygulandığında, ilk ilaçın etkisi hala devam ediyor olabilir. Bu nedenle arınma döneminin süresinin yeterince uzun seçilmesi gereklidir.

Dış kontrollü çalışmaları

Eğer kontrol grubu, başka bir araştırma grubuna aitse, ya da araştırıcının daha önce başka bir yöntem uyguladığı bir grupsa (*historical controls*), bu çalışmaları, dış (*external*) kontrollü çalışmaları olarak adlandırılır (Şekil 7).



Şekil 7
Dış kontrollü klinik çalışma düzeni. Aşağıdaki kalın ok çalışmanın yönünü ve "0" çalışmanın başıldığı anı temsil etmektedir.

Bir hastanede 2000-2003 yılları boyunca portal hipertansiyonu olan 18 hastaya, yeni geliştirilmiş olan bir cerrahi yöntem uygulanıyor. Bu yöntemin sonuçları, 1996-1999 yıllarında daha önceki yöntemin uygulandığı 35 hastadan elde edilen sonuçlarla karşılaştırılıyor.

Örnek 10'daki kontrol grubu, dış kontrol grubudur. Bu çalışma düzeninin en önemli sakıncası, uygulanan yöntem dışındaki etkenlerin etkilerinin neden olduğu farklılıklardır. Örneğin eski yöntemin uygulandığı zamanda antibiyotiklerin daha az etkili olduğu ya da aradan geçen zaman içinde cerrahın deneyim kazanması nedeniyle yeni cerrahi yöntemi daha iyi uygulayabileceği gibi nedenlerle, iki farklı yöntemin karşılaştırılmasında yeni yöntem lehine sonuç elde edilebilir.

NE ZAMAN HANGİ ÇALIŞMA DÜZENİ SEÇİLMELİ?

Her çalışma düzeni her klinik durumda uygulanamaz. Bu nedenle çalışmaya başlamadan önce amaca uygun çalışma düzenlerinden hangisinin seçileceğine doğru karar vermek için her çalışma düzeninin avantaj ve dezavantajlarının iyi bilinmesi gereklidir.

- Yeni tanı ya da tedavi yöntemlerinin etkisinin araştırılması için uygun çalışma düzeni klinik çalışmalarıdır. Klinik çalışmaların içinde de deney ve kontrol gruplarının randomize edildiği çalışmalar en değerli olanlardır. Dış kontrol kullanılan klinik çalışmalarda genellikle yeni yöntem ya da tedavi lehine yorumlanacak sonuçlar verir. Dış kontrollü çalışmalarında, eşzamanlı çalışmalarla göre çok daha fazla oranda yeni tedavinin eski tedavi yöntemlerinden daha iyi olduğu sonucu çıkabilir. Bu nedenle dış kontrollü çalışmalar çok sakıncalıdır.
- Hastalıkların neden, risk etkenleri ve прогнозunu belirlemek amacıyla en uygun çalışmalar prospektif kohort çalışmalarıdır. Bu çalışmaların en önemli zorluğu çok uzun sürmesi ve sonuçlarının geç elde edilmesidir. En meşhur prospektif kohort çalışma, aterosklerotik kalp hastalığının gelişmesindeki etkenlerin ele alındığı Framingham çalışmasıdır. Bu çalışma 1948'de başlamış ve halen devam etmektedir.
- Risk etkeni ile hastalığın ortaya çıkması arasında çok uzun süre geçiyorsa ya da nadir bir hastalıktaki risk etkenleri araştırılıyorsa, retrospektif kohort çalışmaları daha yararlıdır. Ancak bu çalışmalarla asıl ilgilenilen

risk etkeni dışında, hastalığı etkileyen diğer etkenlerin yeteri kadar denetlenmemesi sorunu vardır.

- Retrospektif (*case-control*) çalışmalar da, retrospektif kohort çalışmalar gibi kısa sürede ve düşük maliyetle tamamlanır. Her iki tip çalışmada da amaç aynıdır. Kohort çalışmada çalışma başlangıcındaki tüm denekler sağlıklı, ancak risk etkeni olup olmamasına göre iki gruba ayrılmışken, retrospektif çalışmada hastalıklı ve sağlıklı deneklerden oluşmuş iki gruptaki risk etkenleri retrospektif olarak karşılaştırılır. Retrospektif çalışmalarla kontrol grubunun nasıl seçildiği son derece önemlidir.

ÖRNEK 11

Mesane kanserinde nikotinin etkisinin olup olmadığı araştırılıyor. Kontrol grubu olarak aynı klinikte izlenmiş, mesane kanseri olmayan hastalar (diğer kanserler, astım, peptik ülser vb.) alınıyor. Tüm hastalara sigara içme alışkanlıklar ile ilgili soruları içeren bir anket uygulanıyor. Sonuçta mesane kanseri olanlarda sigara içenlerin oranının daha fazla olduğu ve nikotin ile mesane kanseri arasında bir ilişki olabileceği öne sürülmüyör.

Örnek 11'deki çalışma ilk bakışta iyi düzenlenmiş bir çalışma gibi görünüyor. Ancak kontrol grubuna alınan hastalar arasında astımlı ve peptik ülserli hastaların da yer aldığı ve bu kişilerin hastalıkları nedeniyle hiç sigara içmedikleri ya da daha az sigara içtikleri düşünülürse, sigara içenlerin oranı açısından iki grup arasındaki farkın gerçeği ne kadar yansındığı kuşkuludur. Mesane kanseri ile diğer grup arasında sigara içme oranı açısından gözlenen farklın gerçek nedeni belki, mesane kanseri olanların fazla sigara içmesi değil, kontrol grubundaki hastaların az sigara içmesidir. Bu durumda sigaranın mesane kanserine neden olduğunu öne sürmek hatalı olacaktır. Bu nedenle retrospektif çalışmalarla, bir de tamamen sağlıklı deneklerden oluşan kontrol grubu alınması önerilmektedir.

- Kesitsel çalışmalar özellikle hastalık prevalansını belirlemek amacıyla kullanılır. Tanısal işlemler ve hastalık mekanizmalarını aydınlatmak amacıyla da yapılabilirler. Kısa sürede sonuçlanması nedeniyle yararlı olabilirler. Ama izlem süresinin kısa olmasından kaynaklanan hataları da birlikte taşırlar. Örneğin kesitsel çalışmaların sonuçlarına dayanarak, diyastolik kan basıncının, yaş ilerledikçe, sistolik kan basıncı gibi artmadığı görüşü hakimdir. Ancak Framingham çalışmasının sonuçları bunun böyle olmadığını ortaya koymuştur. Çünkü yaşı ilerledikçe diyastolik kan ba-

sincı artan kişilerde ölüm hızı daha yüksek olduğu için, sağ kalanların ortalaması dijastolik kan basıncı değişmemektedir. Kesitsel çalışmalarla, bu tip önyargılı sonuçlara varılabileceği gözden uzak tutulmamalıdır. Basında sık sık gördüğümüz kamuoyu araştırmaları da kesitsel çalışmalarlardır. Özellikle bu çalışmalardaki sorunlardan biri de çalışma grubuna alınanların bir kısmının çalışmaya katılmamalarıdır.

ÖRNEK 12

Bir araştırmada "Seçimde hangi partide oy vereceksiniz?" sorusuna "A partisi" diyenlerin oranı %45, "B partisi" diyenlerin oranı %33 ve yanıt vermeyenlerin oranı %22 olarak saptanıyor. Buna dayanarak, seçimi A partisinin rahatlıkla kazanacağı öne sürülmüyör.

Örnek 12'deki kamuoyu araştırmasında, eğer B partisine sempati duyan, ama henüz kararsız olan kişiler çoğunluktaysa, yanıt vermeyen %22'lik grup, yanıt veren gruba benzer değildir. Bu grubun hepsi 15 gün sonraki bir araştırmada B partisi lehine fikir değiştirirlerse, A partisinin oranı %45, B partisinin oranı %55 olacaktır.

Sonuç olarak, amaca uygun çalışma düzeninin seçiminin kritik önemi vardır. Bu seçimi yaparken, araştırılacak konunun özellikleri, zaman, para ve personel olanakları gibi etkenler gözönüne alınmalıdır. Aksi takdirde hatalı planlanmış çalışmalardan elde edilecek sonuçların güvenilirliği azalacağı gibi, bu çalışmaların emek ve para kaybı sakıncalarını da birlikte getirir.

bölüm 2

ÖRNEKLEME VE RANDOMİZASYON

Örnekleme

Basit rastgele örnekleme

Sistematik örnekleme

Tabakalı örnekleme

Küme tipi örnekleme

Randomizasyon

Tabakalı analiz

Sonuçların prognostik etkene göre standartizasyonu

Kontrol değişkenlerinin etkilerinin arıtıldığı istatistik modeller

Prognostik etken denetim yöntemlerinin karşılaştırılması

Araştırmancının amacı, bir gruptaki gözlemlerden 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 denenir ve sonuçlar genellenir; yani “bu ilaç hipertansif hastalarda kan basıncını düşürmede etkilidir (ya da değildir)” denir. Çalışmaya alınan deneklerin oluşturduğu gruba örneklem (*sample*), bu grubun temsil ettiği topluma evren (*population*) adı verilir. Yukarıdaki örnekte çalışmaya alınan hipertansif hastalar örneklemi oluşturmaktadırlar. Bu çalışmanın temsil ettiği evren ise tüm hipertansif hastalardır. Bir çalışmanın sonuçları, yalnız çalışma örneklemi 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 örneklem evreni doğru temsil etmesi gereklidir. Bu nedenle örneklemdeki deneklerin evrenden rast-

le (*random*) olarak seçilmesi gereklidir. Örneğin, yukarıdaki çalışma için Ankara'daki tüm üniversite hastanelerinin kayıtlarının alınması yanlıltıcı olabilir, çünkü diğer hastanelerle karşılaşıldığında üniversite hastanelerine tanı ve tedavisinde daha çok sorun olan hastalar yatmış olabilir. Böyle bir örneklemde elde edilecek sonuçlar tüm evrene genellenirse, hata yapılır.

ÖRNEK 13

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 ilaçın etkinliğinin denendiği bir çalışma yapılıyor ve ilaç yararlı bulunuyor.

Örnek 13'teki çalışmadan "Bu ilaç Behçet hastalığında yararlıdır" sonucunu çıkarmadan önce, örneklem evrenden iyi seçilip seçilmediğini araştıralım. Örneklemdeki hastalarda deri tutulumunun hakim olduğu ve eklem ve göz tutulumunun nadir olduğu kestirilebilir. Bu nedenle bu çalışmada örneklemin "Behçet hastaları" evrenini iyi temsil etmediği anlaşılabılır. Bu çalışmanın sonuçlarının genellenebileceği evren "Behçet hastaları" değil "Deri tutulumu olan Behçet hastaları"dır. Çünkü örneklem girecek hastalar "Tüm Behçet hastaları" evreninden rastgele seçilmemişlerdir.

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

ÖRNEK 14

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) 1000 tanesinin (örneklem) seçilip, bu çalışmada kullanılması planlanmış olsun. Hangi dosyaların seçileceği aşağıdaki yöntemlerle belirlenebilir.

Basit rastgele örneklem

Her yatis dosyasının 1'den 12000'e kadar numaralandığını varsayıyalım. Basit rastgele (*random*) örneklemde çeşitli istatistik kitaplarındaki rastgele sayılar tablosundan ya da bilgisayarlardaki örneklem modüllerinden yararlanarak 1000 tane sayı belirlenir ve dosya numarası bu 1000 sayının içi-

de olan hastalar, örneklemme alınır.

Sistematik örneklemme

Bu yöntemde evren genişliği (evrendeki denek sayısı), örneklem genişliği (örneklemdeki denek sayısı) bölünerek elde edilecek rakam kullanılarak, örneklemme alınacak denekler belirlenir. Örneğin $12000 / 1000 = 12$ olduğu için önce 1'den 12'ye kadar olan rakamlardan biri rastgele seçilir. Örneğin bu rakam 5 olsun. Kayıt numarası 5 olan dosya örneklemme seçilir. Daha sonra kayıt numarası $5+12 = 17$ olan, $17+12 = 29$ olan, ... diğer dosyalar seçilerek, 1000 denekli örneklem tamamlanır.

Tabakalı örneklemme

Tamamen rastgele seçim bazen sakıncalı olabilir. Örneğin her yaş grubundan hastalara ait dosyaların örneklemme alınması isteniyorsa, rastgele seçim yapıldığında, bazı yaş gruplarından dosyalar seçilmemiş olabilir. Tabakalı örneklemde, örneklem seçimine başlamadan önce, evren, belirlenmiş özelilikler açısından tabakalara ayrılır ve her tabakadan rastgele seçim yapılır. Yaş dışında, cinsiyet, hastalığın şiddeti ya da evresi, süresi gibi özellikler de tabakalı örneklem için sıklıkla kullanılır. Örneklemde ölçülecek değişkenin özeliliklerine 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 örneklemde bir cinsiyetin, ya da yaş grubunun daha fazla bulunması gibi istenmeyen durumlar ortaya çıkabilir. Tabakalı örneklemde göz önüne alınması gereken bir konu, örneklem seçiminden önce evrene ait bazı özelliklerin bilinmesi gerekliliğidir.

ÖRNEK 15

İkiyüzbin erişkin kişinin yaşadığı bir şehirde anemi prevalansını belirlemek amacıyla bir çalışma yapılıyor. Erkek/kadın oranının "1" olduğu şehirde yaş dağılımının Tablo 1'de "Evren %" sütununda görülen oranlara uygun olduğu daha önceden biliniyor. Çalışmaya rastgele örneklemme yöntemi ile 500 kişi alınıyor. Tablo 1'de "n" sütununda çalışmaya almanın deneklerin sayısı, "Örneklem %" sütununda örneklemdeki deneklerin cinsiyet ve yaş grubu açısından örneklemde ne kadarlık

bir oranı oluşturduğu, "Hb" sütununda altgrupların hemoglobin ortalamaları, "Evren %" sütununda cinsiyet ve yaş grubuna göre evrendeki oranlar görülüyor.

Tablo 1 Evrendeki yaş ve cinsiyet özellikleri dikkate alınmadan oluşturulan bir örneklem evreni iyi temsil etmemesi nedeniyle meydana gelen hata. Seçilen örneklemde evrenin aksine kadınların ve yaşlıların çoğunlukta olması nedeniyle hemoglobin değerleri gerçekte olduğundan daha düşük bulunmuştur.

	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

Tablo 1 incelendiğinde, anemi prevalansı çalışmasına alınan deneklerin %40'ının erkek, %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 örneklemde 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 gerekenen daha düşük olarak bulunmuştur. Bunun nedeni ya tabakalı örneklem yönteminin kullanılmamış olması ya da her tabakaya düşecek denek sayısını belirlerken, evren oranlarının dikkate alınmamış olmasıdır.

Yeniden Örnek 14'e dönersek, hastane yatış kayıtları mevsime göre dört tabakaya ayrılp, daha sonra her mevsimdeki yatış dosyalarından 250'şer tanesi örneklem seçilirse, farklı mevsimlerde yatan hastaların özellikleri açısından rastgele örneklemeye göre evreni daha iyi temsil eden bir örneklem oluşturulmuş olur.

Küme tipi örneklem

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

Örnek 14'te hastalık sıklıklarının mevsimlere göre farklılık göstermesi ne-

deniyle, örneklemde her mevsimden hastaların alınması daha doğru olacaktır. Bunu sağlamamanın bir yolu tabakalı örneklem yöntemiştir. Alternatif olarak küme tipi örneklem de kullanılabilir. Önce 12000 hastane yatış kayıtları ay lara göre 12 kümeye ayrılmış, daha sonra bu 12 kümeye içinden her mevsimden birer ay olmak üzere Ocak, Nisan, Temmuz ve Ekim aylarında yatanlar örneklem alınmaya aday olarak seçilebilir. Her aydaki yatış sayısı aşağı yukarı eşitse, 12000 yatışın 4000'i bu aylarda gerçekleşmiş olmalıdır. Bu sayı hala istediğimiz örneklem genişliği olan 1000'den fazla olduğuna göre bu hastalar arasından dörtte birini seçmemiz gereklidir. Bu kez her aydaki yatış kayıtları haftalara göre dörder kümeye ayrılmış, Ocak ayının ilk haftası, Nisan ayının ikinci haftası, Temmuz ayının üçüncü haftası ve Ekim ayının dördüncü haftasında yatan tüm hastaların dosyaları, örneklem alınabilir. Bu şekilde seçilen örneklemde yaklaşık 1000 denek bulunması umulabilir. Eğer çeşitli altkümelere ayrıldığı halde istenen denek sayısından daha fazla denek kalıyorsa, denek sayısı rastgele ya da tabakalı örneklem yöntemleri ile istenen örneklem genişliğine indirilir.

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

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

RANDOMİZASYON

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 araştırıcının insiyatifinde olmasının önemli sakıncaları vardır. Örneğin yeni ve etkinliği pek de bilinmeyen bir antidepressif ilaçın placeboyla karşılaştırıldığı bir çalışmada araştırmacılar bilinçli ya da bilinçsiz olarak, depresyon bulgularının ağır olduğu bir hastalarda ilacı, hafif olduğu hastalarda placeboyu seçebilir. Böylece ilaç grubunda ağır depresyonlu, placebo grubunda hafif depresyonlu hastalar yer alacağı için elde edilecek sonuçlar ilaç aleyhine olabilir.

Bu nedenle hastaların çalışma gruplarına seçiminin rastgele olmasının

sağlanması gereklidir. Bunu sağlayan yöntemlere randomizasyon adı verilir. Randomizasyon, hastaların tedavi gruplarına seçilmesinde bilinçli ya da bilincsiz 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 olmayı bilir. Bu durumda randomizasyona şu iki yöntemden biri ya da her ikisi birlikte eklenmelidir:

- Tabakalandırma
- Bloklama

ÖRNEK 16

Bir ilaçın etkinliğinin placebo ile karşılaştırılacağı bir çalışmada hastaların ilaç ve placebo gruplarına rastgele olarak seçilmesi planlanıyor. İlacın etkilerinin erkeklerde ve kadınlarda farklı boyutta olacağı beklentiği için, gruplardaki hastaların cinsiyet açısından dağılıminin mutlaka benzer olması isteniyor. Bu nedenle basit randomizasyon yöntemi ile yetinmeyeip, hastaların cinsiyetlerine göre gruplara dengeli olarak dağıtılmışını sağlayacak bir yöntem uygulanmak isteniyor.

Örnek 16'daki 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 secer (bloklar). Örneğin her iki erkek hastadan birini bir gruba, diğerini diğer gruba seçecektir şekilde bir randomizasyon şeması kurar ve kadınları da aynı şekilde secerse, çalışma grupları cinsiyete göre tabakalandırılmış ve ikili gruplar halinde bloklanmış olur. Örneğin ilk erkek hastayı yazı-tura ile A grubuna seçikten sonra gelen 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 bilgisayar-

lardan elde edilebilecek rastgele sayılar tablosundan yararlanarak da randomizasyon yapılabılır. Tablo 2'de cinsiyetin tabaka değişkeni olduğu ve blokların ikişer kişidenoluştugu 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 placebo 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 placebo, dördüncü erkek hastaya ilaç verilecektir.

Tablo 2 Örnek 16'daki çalışma için cinsiyet açısından tabakalandırılmış ve her blokta iki hasta olmak üzere bloklama yapılmış randomizasyon şeması.

Tabaka	Blok	İlaç	Placebo
Erkek	I	+	
	II	+	+
	III	+	+
	IV	+	+
		4	4

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

Birden çok özellik açısından tabakalandırma yapılabılır. Kaç adet tabakalandırma değişkeni kullanılacağının 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.

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

Örnek 17'deki çalışmada uygulanacak tabakalandırmada özen gösterilmesi gereken konu, her tabaka değişkeninin nasıl kategorize edileceğidir. Örnek 16'da "cinsiyet"in iki adet kategoriden oluşacağından kuşku yoktur. Ancak yaşı, vertebra sayısı gibi sayısal değişkenler için tabakaların kategorilerinin araştırmacı tarafından belirlenmesi gerekecektir.

Eğer 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 kategorileri "0-4, 5-9, 10-14, 15-19, ..." gibi beşer yıllık ve "vertebra sayısı" tabakasının kategorileri "1-3, 4-6, 7-9, ..." gibi üçer vertebra olarak belirlenirdiğini ve "skolyoz nedeni"nin ise "doğumsal, edinsel" olarak iki kategoriden olduğunu varsayıyalım. Operasyon bölgesindeki vertebra sayısı "7-9" grubuna giren 6 yaşında (yani 5-9 yaş grubuna giren) doğumsal skolyozlu olan bir hasta desmopressin grubuna alındığında, placebo grubuna da vertebra sayısı ve yaşı benzer olan doğumsal skolyozlu bir hastanın alınması gerekecektir. Aksi takdirde gruplar, prognostik etkenler açısından dengeli olmayaç 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 müdahaleye yapılacak bir hastayla 10 vertebraya müdahaleye 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 yapılmasıdan 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’in altında - üstünde” olarak iki kategoriye ayrılması yeterli olacaksız “20-24, 25-29, 30-34, ...” gibi dar kategoriler oluşturmaya gerek yoktur.

Tabakaların kategorilerinin eşit genişliklerde olması gerekmekz. Örneğin özellikle ilk yaşta ve 60 yaşından sonra daha ağır seyreden ve bu nedenle ilaç yanıt açısından bu yaş grublarında 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 gruppala ayrılmışıdır.

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).

Tablo 3 Örnek 17’deki çalışmada yaşı, vertebra sayısı ve skolyoz nedeni açısından tabakalandırma şemasına göre desmopressin ve placebo grubuna alınan hasta sayıları.

Yaş	Vertebra sayısı	Skolyoz nedeni	Desmopressin	Placebo
<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

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

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

Acaba beta-bloker grubunda infarktüs oranının düşük olmasının nedeni ilaçın 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 erkeklerde ve kadınlarda tedavi gruplarındaki ayrı ayrı incelenmesi grupların cinsiyet açısından dengesiz olmasının neden olduğu yorumlama güçlüğü ortadan kaldıracaktır. Gerçekten de hem erkeklerde ($p=0.41$) (Tablo 8), hem de kadınlarda ($p=0.46$) (Tablo 9) hangi ilaç kullanılırsa kullanılsın, прогнозun farklı olmadığı görülmektedir.

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

		Miyokard İnfarktüsü		Toplam
	+	-		
Beta-bloker	8	2		10
Kalsiyum antagonistisi	27	3		30
Toplam	35	5		40

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

		Miyokard İnfarktüsü		Toplam
		+	-	
Beta-bloker	13	17	30	
	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 olmayı bilir.
- Gruplararası karşılaştırma sayısı artacağı için “istatistiksel önemlilik enflasyonu” riski vardır.

Yeri gelmişken çalışmanın gücü ya da istatistik analizin gücünü ne anlam geldiğini açıklamakta yarar var. Örneğin iki ilaca yanıtın incelendiği bir araştırmada, iki grupta tedaviye yanıt oranı arasında fark olduğu varsayımları araştırmada toplanan verilere dayanarak test edilir ve analiz sonucunda genellikle herkesin çok aşina olduğu bir p değeri elde edilir. P değeri, kabaca iki grup arasında gerçekte fark olmadığı halde, şans eseri olarak fark varmış gibi sonuç elde edilmiş olması olasılığıdır. Örneğin bir ilaca yant 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 yapılsa, 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 saptanabilecegi anlamına gelir.

Bu araştırmayı analizini 6 tane çapraz tablo ile yapabildik (Tablo 4-9). MI 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 azaldı. Eğer araştırmacıın başlangıcında cinsiyetin prognoz üzerine etkisi olabileceğü öngörülse ve buna göre hastalar tabakalandırılarak randomize edilseydi, 80 hastayla yapılmış olan bu araştırmadaki istatistik analizinin gücüne, aşağı yukarı 40 hasta ile yapılacak bir çalışmaya ulaşılabilirdi.

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 enfasyonuna neden olacağı için de sakıncalıdır. İstatistiksel önemlilik enfasyonu, 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 tartışılar 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 ağırlıklandırarak kullanılmıştır. Ağırlıkların seçimi, analizin amacına göre değişir.

Standardizasyonun nasıl yapıldığını şu örnekte inceleyelim:

ÖRNEK 19

Hipertansiyonda A ve B ilaçlarının placebo 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 yapılmasıının doğru olmayacağı düşünülmüştür.

Tablo 10 Örnek 19'daki çalışmada grplardaki 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 19'daki çalışmada grplar 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 olmayacağından emin olmak gerekmektedir.

Tablo 11 Örnek 19'daki çalışmada grplarda 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

Grplar cinsiyet açısından dengeli olmadığı için, tedaviye yanıtın nasıl olacağını kestirmek için sonuçlar standardize edilmelidir. A ilacı alan hastalar da gözlenen ortalama kan basıncı düşmesi 23.9 mmHg'dır. 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 dağılımin aynı olduğu için, yanı yarı yarına erkek ve kadın olduğu için standardize ortalama, gözlenen ortalamadan farklı değildir. B ilacı verilen grupta gözlenen ortalama kan basıncı düşmesi 34.0 mmHg'dır. 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'dır. Bu üç ortalama, artık cinsiyetin etkisinden arıtlımıştır, çünkü her üçü de %50'si erkek, %50'si kadın olan hipotetik popülasyonlarda gözlenmesi beklenen değerlerdir.

Kontrol değişkenlerinin etkilerinin arıtıldığı istatistik modeller

Gruplar arasında dengesizlik olduğu zaman kullanılabilecek bir yöntemin tabakalı analiz olduğundan ve bu yöntemin sakıncalarından yukarıda söz edildi. Dikkat ederseniz, tabakalı analizde prognostik etkene göre her altgrubun analizini ayrı ayrı yaparken, analize aldığımız hasta sayısı çalışmaya aldığımız hasta sayısından daha az oluyor ve buna bağlı olarak çalışmanın gücü azalıyordu. Tabakalı analizin bu sakıncalarını ortadan kaldırın ve sonuç üzerine prognostik etkenlerin (kontrol değişkenlerinin) etkilerinin arıtıldığı 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 başıta gözlenen farkın iki gruptaki hastaların cinsiyete göre dağılımları arasındaki farktan kaynaklandığı 6 farklı çapraz tablonun istatistiksel analizi ile gösterilmiştir. Ancak bu sonuca ulaşabilmek için çalışmanın gücünün azalması ve istatistiksel önemlilik enflasyonu riskini göze almıştık. Bu araştırmada alternatif olarak log-lineer analiz yapıldığında iki tedavi grubundaki hastaların cinsiyet dağılımının benzer olmadığı ($p<0.0001$), прогнозun iki cinste benzer olmadığı ($p=0.0001$), ancak tedavi ile прогноз arasında ilişki olmadığı ($p=0.88$) saptanmıştır. Burada üzerinde önemle durulması gereken nokta tedavi ile прогноз arasında ilişki olmadığını söylemek, cinsiyetin etkisinin arıtlımsız olmasıdır.

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

Sonradan-tabakalandırma ile grup karşılaştırmaları sayılarındaki artmaya bağlı "istatistiksel önemlilik enflasyonu", bu istatistik modellerle önlenebilir. Bu yöntemlerle, hem tedavinin etkisi prognostik etkenlerden arıtlımsız, 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 yanılığı 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 sınırlı olan yargıla-

Tablo 12 Kontrol değişkenlerinin etkilerinin arthedüğü istatistik modeller.

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

* Yanıt kriteri "yanıt var-yok" ya da "öldü-yasıyor" gibi iki sonuçtan ibaretsen.

** Yanıt kriteri "yanıt tam-kısmı-yok" ya da "haftı-orta-agır" gibi ikiden daha çok sonuç içeriyeorsa.

rin doğruluğu artabilir, ya da bazen tersine farklı 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. Herseye 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.

Araştırmmanın amacı, çalışmaya alınan hastalardaki sonuçların tüm hastalara genelleştirilebilmesidir. Bu yazında anlatılan randomizasyon ve prognostik etken denetim çabalarının nedeni budur.

bölüm 3

DEĞİŞKENLER VE ÖZELLİKLERİ

Değişkenlerin ölçümü

Nominal ölçüm

Ordinal ölçüm

Sayısal ölçüm

“Önemli rakam” sayısı

Değişkenlerin dağılım özelliği

Dağılım şekli ölçütleri

Verilerin dağılım özelliği niçin önemlidir?

SPSS'te normal dağılım testleri nasıl yapılır?

İstatistikte değişken (*variable*) terimi, deneklere ait özellikler anlamında kullanılır. Örneğin iki gruptan birine A ilaçı, diğerine B ilacının verildiği ve hastaların ilaca verdikleri yanıtın karşılaştırıldığı bir klinik çalışma düşünelim. Bu çalışmada hastaların yaşı, bir değişkendir. Aynı şekilde cinsiyet, boy, ağırlık, kan basıncı, hangi ilaç grubuna girdikleri, tedavi öncesi kan basıncı vb. hepsi birer değişkendir. Yani, her hastaya ait veri giriş formundaki her bilgi bir değişkendir.

DEĞİŞKENLERİN ÖLÇÜMÜ

İstatistik analize başlamadan önce ilk yapılacak şey, değişkenlerin nasıl ölçüldüğünün belirlenmesidir. Bu sorunun yanıtı “sfigmomanometre, Sahli yöntemi, termometre, otoanalizer, ...” gibi cihaz ya da yöntemler ya da “santimetre, yıl, IU/mL” gibi birimler olabilir; ancak istatistik biliminde ölçüm'den kastedilen, bu değildir. İstatistikte ölçüm denildiği zaman anlaşılan, değişkenin alabileceği değerlerle ilgili kısıtlamalardır. Örneğin bir kadının gebelik sayısı 5.5 olamaz, ama yaşı 25.64567 yıl olabilir. Ya da yeni doğan bir bebeğin APGAR skoru 0 ile 10 arasındaki tüm tam sayılar olabilir, ama 8.4 ya da 15 olamaz. Değişkenlerin alabileceği değerlerin neler olabileceği, neler

olamayacağı, yani nasıl ölçüldüğünü belirlemek, yapılacak istatistik analizin seçimi için çok önemlidir. Ölçüm özelliklerine göre değişkenleri başlıca üç gruba ayıralım: nominal, ordinal ve sayısal değişkenler.

Nominal ölçüm

Nominal bir değişkende ölçüm düzeyleri arasında bir sıralama ya da uzaklık-yakınlık gibi belirli bir mesafe yoktur.

Ölçüm düzeyleri (*levels of measurement*), değişkenin alabileceği değerler anlamındadır. Örneğin tip I diabetes mellitus (DM), tip II DM ve normal sağlıklı deneklerin bulunduğu bir çalışmayı ele alalım. Her denek için bir veri giriş formu hazırlanmışsa, formda "HASTANIN AİT OLDUĞU GRUP" bölümüne üç farklı değer girilebilir. O halde bu değişkenin 3 ölçüm düzeyi vardır. Normaller "0", tip I DM "1" ve tip II DM "2" olarak kodlanmış olsun. Burada 0, 1 ve 2 aslında rakam olarak bir anlam ifade etmemektedir. Deneklerin ait oldukları gruplar "0, 1, 2" ile değil de "9, 1, 2" ya da "NOR, BİR, İKİ" diye de kodlanabilir. Nominal değişken değerleri, rakam olarak değil, ad olarak anlam ifade ederler. Başka bir çalışmada grup değişkeni sırasıyla "normal, kronik aktif hepatit, siroz, hepatoselüler kanser ve biliyer siroz" gruplarını temsil eden 1, 2, 3, 4 ve 5 değerlerini alan nominal bir değişken olabilir. Değişken değerleri "1, 2, 3, 4, 5" yerine "NOR, KAH, SİR, CA, BİL" olsa da farketmez. Nominal değişkenlerin aldığı değerlerin mutlaka rakam olması gerekmekz. Örneğin "İNFEKSİYON ETKENİ" adlı bir değişken "viral, bakteriyel, fungal, protozoal" gibi değerleri alabilen bir nominal değişkendir. Ama pratik amaçlar için bu tip nominal değişkenlerin alabileceği tüm değerleri kodlayarak, rakama dönüştürmekte yarar vardır.

İki düzeyli, yani yalnız iki değer alabilen nominal değişkenlere ikili değişkenler (*dichotomous variable*) adı verilir. Örneğin "CİNSİYET", erkek ve kadın olarak iki düzeyi olan bir ikili değişkendir. Aynı şekilde "evet, hayır" olarak iki değer alabilen bir değişken de ikili değişkendir.

Ordinal ölçüm

Ordinal bir değişkende ölçüm düzeyleri arasında bir sıralama vardır, ama düzeyler arasındaki mesafeler belirli değildir.

Örneğin tümörlü hastaların evresinin girildiği "EVRE" adlı bir değişken olsun. Bu değişkene evre 1'den 4'e karşılık gelmek üzere, 1, 2, 3 ve 4 değerleri girilebilir. Bu değerler belirli bir sıra ifade etmektedir. Örneğin "evre 3, evre 2'den daha ileri evredir", "evre 1 en iyi, evre 4 en kötü evredir" vb. Ama değişkenin düzeyleri arasındaki mesafeler belirli değildir. Örneğin matematik işlem yapıldığında $2-1 = 4-3 = 1$ doğrudur, ama "evre 2, evre 1'den ne ka-

dar ileriye, evre 4 de evre 3'den o kadar ileridir" denemez.

Ordinal değişken değerleri yalnızca " $>$ " ve " $<$ " işlemleri için sayı gibi değerlendirilir; bunlar dışındaki matematik işlemler uygulanamaz.

Yukarıdaki örnekte denekler "1, 2, 3, 4" yerine "1, 4, 5, 9" ya da "8, 6, 4, 2" şeklinde de kodlansa, sıralama değişimeyeceği için farketmez. Ama "1, 0, 2, 3" şeklinde kodlamak hatalı olur, çünkü bu durumda ordinal değişkenin ölçüm düzeyleri arasında sıralama olması koşuluna uyulmamış olur. Ordinal değişken değerleri arasında çıkartma işlemi yapılamayacağı için, sıralamayı korumak koşuluyla istenilen rakamlarla kodlanabilir.

Nominal ve ordinal ölçümlle belirtilen değişkenlere kategorik değişkenler adı da verilmektedir.

Sayısal ölçüm

Bir değişkenin aldığı değerler, nominal ya da ordinal değişkenlerde olduğu gibi araştırıcı tarafından belirlenmiş kodlar değil de gerçek rakamlar, o değişkenin sayısal ölçüm skalasında ölçüldüğü söylenebilir.

Sayısal ölçümlle belirlenen değişkende, değişken düzeyleri arasında hem sıralama, hem de belirli bir mesafe vardır. Sayısal değişken değerlerine reel sayılarla uygulanan her türlü matematik işlem uygulanabilir.

Sayısal değişkenler iki şekilde olabilir: kesikli ya da sayımla belirtilen sayısal değişken (*interval variable*) ve sürekli ya da ölçümlle belirtilen sayısal değişken (*continuous variable*). Eğer bir sayısal değişken her değeri alamıyor, ancak belirli değerleri alabiliyorsa, o değişkene kesikli değişken adı verilir. Sayımla belirtilen değerleri alabilen değişkenler kesiklidir. Her türlü değeri alabilen değişkenler ise, sürekli ya da ölçümlle belirtilen değişkenlerdir. Örneğin, bir ailedeki çocuk sayısı 1, 2, 3, ... gibi yalnızca tam sayı şeklindeki değerleri alabilir. Bu nedenle "ÇOCUK SAYISI" kesikli ya da sayımla belirtilen bir değişkendir. Bir insanın yaşı ise genellikle tam sayılarla belirtile bile, teorik olarak 24.5 yıl, 34.56 gün gibi her türlü değeri alabilir, bu nedenle "YAŞ" sürekli bir değişkendir.

Sayısal değişkenler, sınıflandırılarak ordinal değişkenlere dönüştürülebilir. Örneğin kalsiyum düzeyi sürekli değişkendir. Eğer kalsiyum düzeyinin rakam olarak değerinden çok, "normalden düşük", "normal" ya da "normalden yüksek" olması önemliyse, "8.9'dan düşük", "8.9-10.1 arasında" ve "10.1'den yüksek" olarak yalnız üç değer alabilen bir ordinal değişkene dönüştürübiliriz. Hastaların yaşları da "20'den küçük", "20-34", "35-50", "50'den büyük" gibi sınıflandırılarak, ordinal değişkene dönüştürülebilir.

Çok gerekmekle, sayısal değişkenleri ordinal değişkenlere dönüştürmek uygun değildir. İstatistik analiz sırasında hataya yol açmaz, ama daha

az bilgi veren yöntemler kullanılmasını gerekli kılabılır. Bu nedenle çalışmalar sırasında verileri toplarken, deneklere ait sayısal değişkenler sınıflandırılmadan, gerçek değerleri ile kaydedilmelidir.

“Önemli rakam” sayısı

Bir çalışmada “hastanın yaşı”, “Hemoglobin (Hb) değeri” ve “beyaz kürdeğeri” gibi üç değişken olduğunu varsayıyalım. Hastanın yaşı “yıl” olarak, yani 23, 64 gibi iki tane rakam içeren tam sayılarla belirtilmişse, yani hastalarla yaşları daha ayrıntılı olarak sorulmamış ya da sorulduğu halde kaydedilmemişse, bu değişkendeki değerlerin önemli rakam (*significant digit*) sayısı ikidir. Hb değeri 12.3 g/dL gibi 0.1 g/dL duyarlılıkla ölçülebiliyorsa, bu değişkendeki değerlerin “önemli rakam” sayısı üçtür. BK değeri 7400/mm³ ya da 3500/mm³ gibi 100/mm³ duyarlılıkla sayılabiliriyorsa, BK değerlerinin “önemli rakam” sayısı ikidir.

Yani bir sayıdaki önemli rakam sayısı o sayının ne kadar duyarlı bir şekilde ölçüldüğü ile ilgili olarak belirlenir. Bir rakam, ondalık ayırcı noktanın sağında ya da solunda da olsa, Hb örneğinde olduğu gibi önemli rakam olabilir. Öte yandan BK örneğinde olduğu gibi noktanın solunda bulunan tüm rakamların önemli rakam olması gerekmek.

Sayırlarla matematik işlem yaparken, önemli rakam sayılarına dikkat etmek gereklidir. Örneğin hastaların yaş ortalamalarını verirken, bir hastanın yaşındaki önemli rakam sayısından daha fazla önemli rakam içeren bir sayı kullanılması hatalıdır. Hastaların yaşı 23, 64 gibi iki adet önemli rakamı olan tam sayılarla belirtilmişse, yaş ortalaması 54.2, 54.23 gibi ikiden fazla önemli rakam içeren sayılarla verilmemelidir. Aynı şekilde 0.1 gr/dL duyarlılıkla ölçülen Hb değerlerinin ortalaması 13.34 gr/dL, 100/mm³ duyarlılıkla ölçülen BK değerlerinin ortalaması 6530/mm³ ya da 6533/mm³ gibi sayılarla belirtilemez.

Kural olarak toplama, çarpma, karekök alma vb. matematik işlemlerinin sonucundaki “önemli rakam” sayısı, işlem yapılan sayılardaki en düşük “önemli rakam” sayısından fazla olamaz.

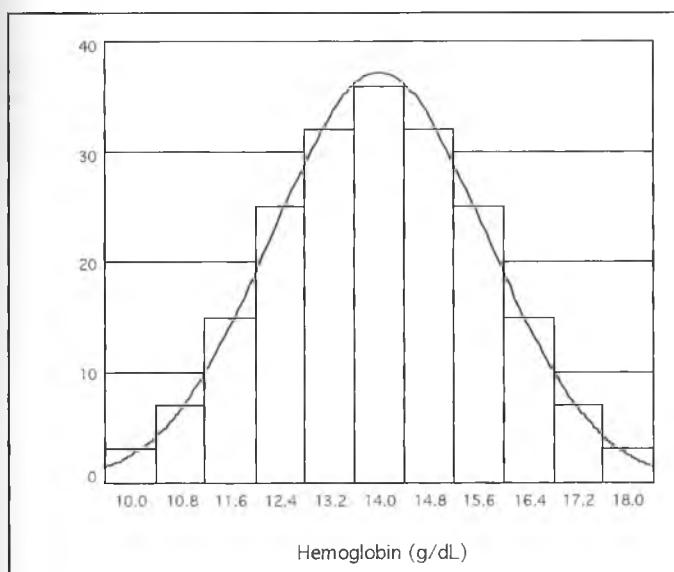
Örneğin BK değeri bir çalışmaya alınan 40 denekten 20'sinde 100/mm³, 20'sinde 10/mm³ duyarlılıkla ölçülmüşse, yani ilkinde 7200/mm³ gibi daha kaba, ikincisinde 8640/mm³ gibi daha duyarlı bir şekilde belirtilmişse, önemli rakam sayısı birinci durumda iki, ikinci durumda üçtür. Bu 40 kişinin BK ortalamalarını hesaplarken toplama ve bölme işlemleri yapacağımız için, en düşük önemli rakam sayısından (yani üç yerine ikiden) daha fazla önemli rakam içerecek kadar duyarlı bir ortalama belirtmemeyiz. Hesapladığımız BK ortalaması 5632.32/mm³ olsa bile, bunu 5632/mm³ ya da 5630/mm³ şeklinde belirtmemeyiz. Doğru gösterim şekli 5600/mm³ olmalıdır.

DEĞİŞKENLERİN DAĞILIM ÖZELLİĞİ

ÖRNEK 20

Rastgele 200 kişinin hemoglobin (Hb) düzeyleri ölçülüp, düşükten yükseğe doğru sıralanıyor. En düşük değerin 9.8 g/dL, en yüksek değerin 18.2 g/dL ve ortalama değerin 14.0 g/dL olduğu saptanıyor.

Örnek 20'deki 200 kişinin Hb değerlerinin çoğunun ortalamaya yakın olduğunu, düşük ve yüksek değerlere gidildikçe, sayının azaldığını görürüz. Örneğin Hb'i 13.6-14.4 g / dL arasında olan otuzaltı kişi varken, 12.0-12.8 g / dL arasında olan yirmibeş, 10.4-11.2 g / dL arasında olan yedi ve 10.4 g / dL'den düşük olan üç kişi olduğunu görebiliriz. Yatay eksende Hb değeri, dikey eksende de denek sayısı olmak koşuluyla çubuk grafik çizilirse (yani örneğin Hb değeri 13.6-14.4 g / dL olan otuzaltı denek olduğunu varsayırsak, yatay eksendeki 14.0 g / dL'den yukarı dikey eksenin 36 birimi uzunlığında çubuk çizilip, tüm Hb değerleri için bu şekilde grafik tamamlandığında), çubukların tepe noktalarını birleştiren eğri, Hb değerlerinin dağılım eğrisidir (Şekil 1).



Şekil 1
Normal
dağılım eğrisi.

Doğadaki ve toplumdaki çok şeyin dağılımı buna benzer. Normal dağılım adı verilen bu dağılım şekli simetriktir. Normal dağılım eğrisi, çan şeklinde olduğu için çan eğrisi olarak da adlandırılmıştır.

Yukarıdaki örnekte Hb'i 12.0-12.8 g/dL arasında olan yirmibeş kişi olduğunu varsayımyışık. Bu aralığın ortasında bulunan 12.4 g/dL, ortalamadan (14.0 g/dL) 1.6 g/dL azdır. Dağılım normalse, Hb değeri ortalamadan 1.6 g/dL fazla olan (yani orta değeri 15.6 g/dL, değer aralığı 15.2-16.0 g/dL olan) kişilerin sayısının da yirmibeşe yakın olması gereklidir.

Normal dağılım simetrik olduğu için normal dağılım gösteren değişkenlerin ortalama, ortanca ve modları eşittir.

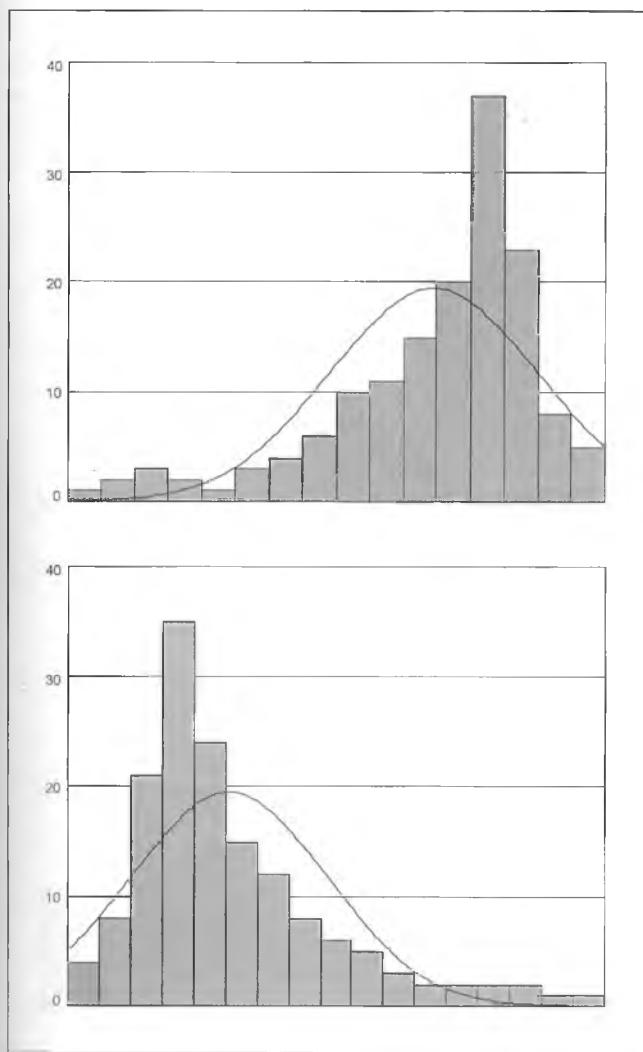
İstatistikte dağılımin normal olup olmadığı belirlenmesi çok önemlidir. Çünkü farklı dağılım gösteren verilere uygulanacak tanımlayıcı ve analitik istatistik yöntemleri de farklıdır.

Dağılım şekli ölçütleri

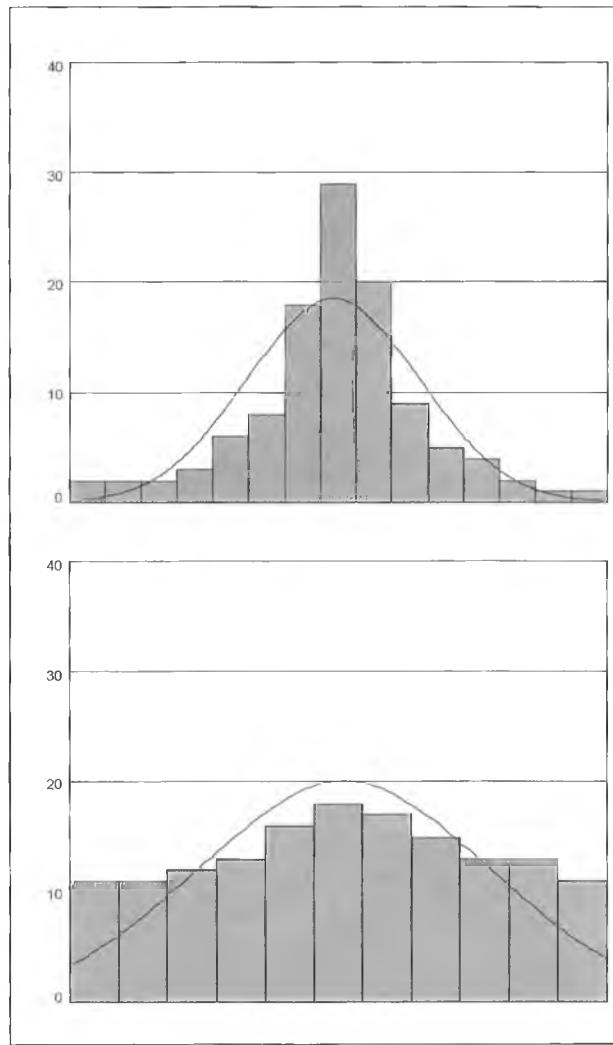
Eğer bir dağılım simetrik değilse, dağılım eğrisinde çan eğrisinin tepe noktası ortada değil, sağa ve sola kaymış olacaktır. Başka bir deyişle, denekler hangi tarafta daha çoksa, eğrinin öteki tarafı basıklaşacaktır.

Denekler ortalamadan daha büyük değerlerde toplanıyorsa, negatif çarpık ya da sola çarpık (*negatively skewed, skewed to the left*), küçük değerlerde toplanıyorsa, pozitif çarpık ya da sağa çarpık (*positively skewed, skewed to the right*) dağılımdan sözedilir (Şekil 2).

Eğrinin tepesinin sivriliğinin şekli de basıklık (*kurtosis*) ile belirtilir. Tepe si sivriyse, leptokurtik, tepesi basıksa, platikurtik dağılımdan söz edilir (Şekil 3). Çarpıklık ve basıklık dereceleri de histogramlardan anlaşılabılır, ama bu özellikleri temsil eden indekslere de bilmek yararlı olabilir. Normal dağılım için çarpıklık ve basıklık indeksleri 0'dır. Pozitif değerler, çarpıklık için pozitif çarpıklığı, basıklık için leptokurtik olduğunu gösterir.



Şekil 2
Sağa (pozitif)
(üstteki şekil) ve
sola (negatif)
(alttaki şekil)
çarpık dağılım eğrisi.



Şekil 3
Leptokurtik
(üstteki şekil) ve
platykurtik (alttaki
şekil) dağılım eğrisi.

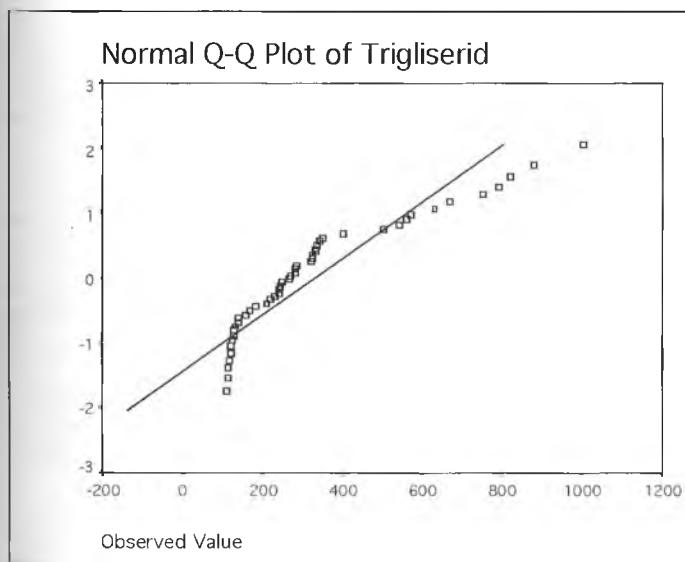
Verilerin dağılım özelliği niçin önemlidir?

İstatistikteki en önemli konulardan biri dağılım özelliğidir. Parametrik testlerin tümünün uygulanabilmesi için gereken varsayımların başında verilerin dağılımının normal olması gelir. Dağılımin normal olup olmadığı grafik ve istatistik analiz yöntemleri ile araştırılır.

Değerlerin histogramını çizerek dağılımın normalden sapıp sapmadığını kabaca anlamanın olası olduğunu görmüştük. Bir diğer grafik yöntem de normal olasılık grafiği (*normal probability plot*) adı verilen grafiğin çizilmesidir. Bu grafikte bir eksen gözlenen değerler, diğer eksen ise normal dağılım olduğu varsayımlına göre beklenen değerler eksenidir.

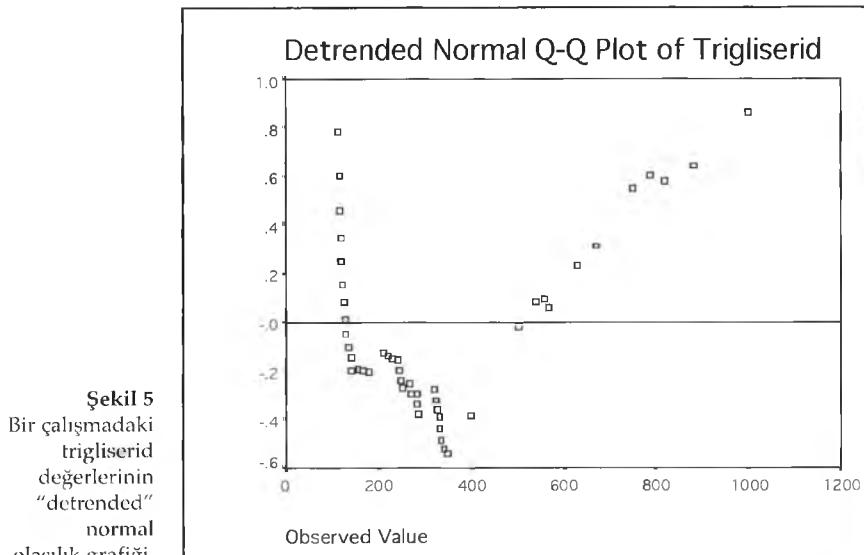
Veriler normal dağılım gösteriyorsa, noktaların bir doğru üzerinde yer alması ya da etrafında belirli bir pattern göstermeden dağılması gereklidir.

Şekil 4'de bir çalışmada hastaların trigliserid değerlerine ait normal olasılık grafiği örneği görülmektedir. Noktalar bir doğru oluşturmamaktadır ve eğrisel bir pattern göstermektedir. Bu nedenle normal dağılım göstermediği rahatlıkla söylenebilir.



Şekil 4
Bir çalışmada
trigliserid
değerlerinin
normal olasılık
grafiği.

Her değerin gözlenen değeri ile normal dağılım olduğu takdirde beklenen değeri arasındaki farkların noktalandığı *detrended* normal olasılık grafiği'nde ve beklenen, noktaların dikey eksende "0"dan çizilen yatay çizgi etrafında dağılmış olarak yer almalarıdır. Böyle değilse, yine normal dağılım olmadığı varsayılabılır. Şekil 5'te trigliserid değerlerine ait *detrended* normal olasılık örneği görülmektedir. Bu grafikten de dağılımın normal olmadığı söylenebilir.



Şekil 5
Bir çalışmada triglycerid değerlerinin "detrended" normal olasılık grafiği.

Grafik yöntemlerden dağılımin normal olup olmadığı hakkında fikir edilebilir. Ama bu izlenimin istatistik yöntemlerle test edilmesi gerekir. Shapiro-Wilks' ve Lilliefors testleri bu amaç için sıkılıkla kullanılan testlerdir. Bu testlerde p değeri <0.05 ise, dağılımin normal olmadığı sonucuna varılır. Ancak tamamen normal dağılıma uygun verilerle çalışmak hemen hemen imkansızdır. Özellikle denek sayısı fazlaysa, genellikle tüm normal dağılım testleri ile " $p<0.05$ " bulunur. Bu nedenle bu durumlarda *detrended* normal olasılık grafiği gibi grafik değerlendirmelerde normalden çok ciddi sapmalar yoksa (özellikle belli bir pattern çizerek), dağılımin normal olduğu varsayılmalıdır.

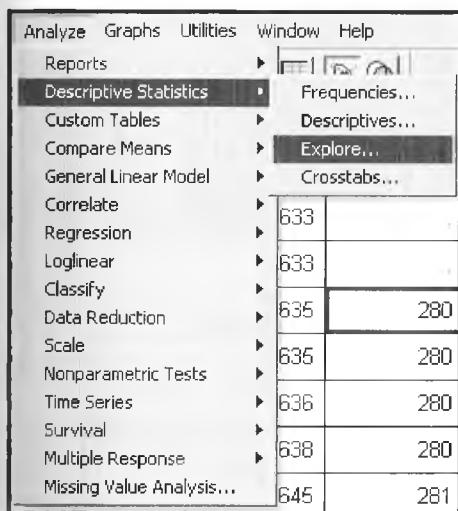
SPSS'TE NORMAL DAĞILIM TESTLERİ NASIL YAPILIR?

SPSS'te dağılımin normal olup olmadığı çeşitli yöntemlerle test edilebilir:

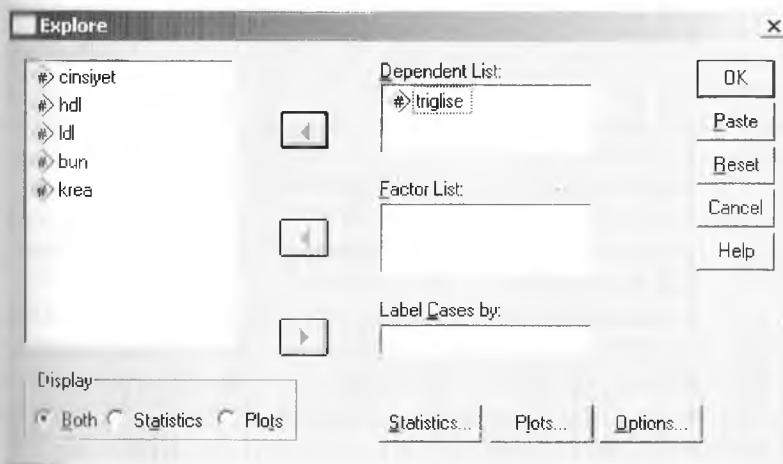
1. "Analyze >> Descriptive Statistics >> Frequencies" diyalog kutusundan ulaşılan Charts diyalog kutusu yoluyla histogram çizerek dağılımda normalden sapmalar olup olmadığı hakkında fikir edinilebilir. Bölüm 7'de ayrıntılı olarak anlatılmıştır.

2. "Analyze >> Descriptive Statistics >> Explore" seçimi ile açılan Explore diyalog kutusundaki Plots seçeneğinde *Normality plots with test* ile normal

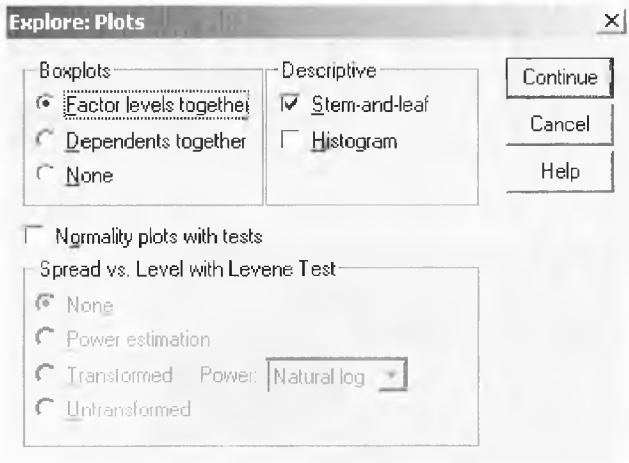
olasılık ve *detrended* normal olasılık grafikleri ve normal dağılım testleri yapılır (Şekil 6, 7 ve 8). Dağılımı test edilecek değişken *Dependent List*'e alınmalıdır. Eğer tüm deneklerdeki ölçümlerin dağılımı test edileceksse *Factor List* boş bırakılmalıdır. Altgruplarda ölçümlerin dağılımı ayrı ayrı test edileceksse, altgrupları belirleyen değişken *Factor List*'e seçilmelidir.



Şekil 6
SPSS'te "Analyze > Descriptive Statistics > Explore" seçimi.



Şekil 7
SPSS'te
"Explore"
diyalog
kutusu.



Şekil 8
SPSS'te "Explore:
Plots" diyalog
kutusu.

Şekil 9
Triglicerid değerlerinin
normal dağılım gösterip
göstermediğinin "Analyze >>
Descriptive Statistics >>
Explore" menüsünden
yapılan istatistik
analiz sonuçları.

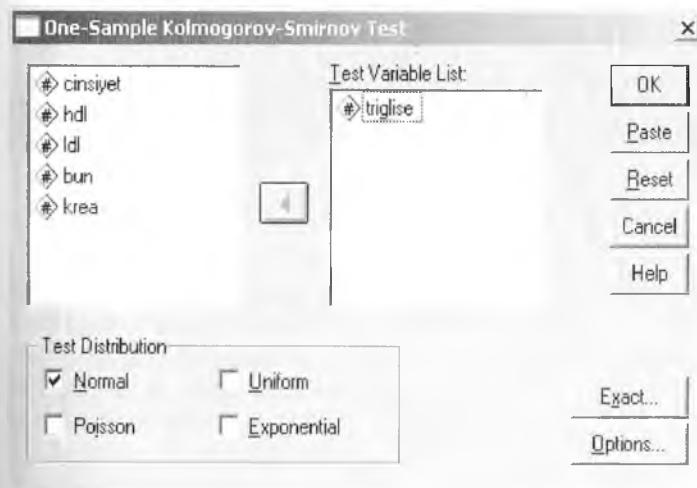
Tests of Normality			
	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
Triglycerid	.116	112	.000

^a Lilliefors Significance Correction

Denek sayısı 30'dan azsa, Shapiro-Wilks ve Lilliefors testleri uygulanır. Daha fazla denek varsa, Shapiro-Wilks testinin hatalı sonuç verme olasılığı olduğu için yalnız Lilliefors testi yapılır. Bu testlerdeki p değerleri <0.05 ise dağılımin normal olmadığı anlaşılır. Şekil 9'da triglycerid değerlerinin normal dağılmadığı istatistik analiz yöntemleri ile de doğrulanmıştır.

P değeri *significance* olarak gösterilmiştir. SPSS'te p değeri genellikle noktanın sağında dört rakama kadar belirtilir. Burada ".000" olarak gösterilen p değeri " <0.001 " olarak alınabilir.

3. "Analyze >> Nonparametric tests >> 1-Sample K-S" seçimi ile One-Sample Kolmogorov-Smirnov Test (Tek örneklem Kolmogorov-Smirnov testi) diyalog kutusu açılır (Şekil 10). Bu pencerede *Test Distribution* (Test dağılımı) menüsünden *Normal* seçeneği seçildiğinde yapılan testte p <0.05 ise dağılımin normal olmadığı anlaşılır.



Şekil 10
SPSS'te "One-Sample Kolmogorov-Smirnov Test" diyalog penceresi.

Dağılıminın normal olup olmadığı kararını verirken grafik yöntemler, istatistik analiz yöntemlerinden daha yararlı olabilir.

Şekil 11'de trigliserid değerlerinin dağılıminın normallığının Kolmogorov-Smirnov testi ile yapılan analizinin sonuçları gösterilmiştir. *Asymp. Sig. (2-tailed)* olarak verilen p değeri >0.05 olduğu için dağılımin normal olduğu sonucu çıkmaktadır.

One-Sample Kolmogorov-Smirnov Test		
N		Trigliserid
Normal Parameters	Mean	112
	Std. Deviation	597.21
Most Extreme Differences	Absolute	48.05
	Positive	.116
	Negative	.116
Kolmogorov-Smirnov Z		-.090
Asymp. Sig. (2-tailed)		1.233
		.096

Şekil 11
Trigliserid değerlerinin normal dağılım gösterip göstermediğinin "Analyze >> Nonparametric Tests >> 1-Sample K-S" menüsünden yapılan istatistik analiz sonuçları.

bölüm 4

VERİTABANININ HAZIRLANMASI

Nominal ve ordinal verilerin kodlanması

Değişkenleri adlandıırken dikkat edilecek noktalar

Veri girişinde hata kontrolü

SPSS'te veri ve dosya işlemleri

Veritabanı nasıl hazırlanır?

Değişkenlerin tanımlanması

Değişken adı

Değişken tipi

Değişken ve değer etiketleri

Kayıp değerler

Sütun formatı

Ölçüm tipi

"Template" nedir, ne işe yarar?

Veri girişi ve düzenlenmesi (edit) nasıl yapılır?

Antihipertansif verilen gruptaki ortalama kan basıncı, placebo verilen grubun değerinden daha mı düşük bulunacak, ya da koroner arter hastalığı grubunda sigara içme oranı ve ortalama LDL düzeyi, normal gruptan daha mı yüksek? Klinik çalışmaların çoğunda merakla beklenen sonuç, deney ve kontrol grupları arasında fark olup olmadığıdır. Bu sonuçları verecek istatistik analizlere başlamadan önce, veriler bazı özellikleri açısından gözden geçirilmezse, önemli hatalar yapılabılır. Bu nedenle her türlü çalışmanın istatistik analizinde şu sırayı izlemek gereklidir:

1. Nominal ve ordinal verilerin kodlanması.
2. Verilerin veritabanına (bilgisayara ya da kağıda) girilmesi.
3. Veri girişi sırasında yapılan hatalar açısından veritabanının gözden geçirilmesi.
4. Verilerin tanımlayıcı istatistiğinin yapılması.
5. Gerekirse, verilere dönüşüm uygulanması.
6. Çalışma amacına ve verilerin dağılım özelliklerine uygun istatistik analizi

lerin yapılması.

7. İstatistik analiz sonuçlarının klinik açıdan yorumlanması ve olası değişimlerin etkileşimlerinin saptanması.
8. Gerekirse, etkileşimleri arıtmak için istatistik analiz yöntemleri uygulanması.
9. Çalışma raporunda belirtilecek istatistik sonuçların seçilmesi ve raporun yazılması.

Bu sırayı izlerken yapılacak acelecilik, ciddi hatalara yol açabilir.

NOMİNAL VE ORDİNAL VERİLERİN KODLANMASI

Nominal değişkenleri kodlarken, kontrol ya da placebo grubunun en küçük ya da en büyük rakamla kodlanması yararlıdır. Örneğin “koroner arter hastalığı, kardiyomiopati, kapak hastalığı ve normal kontrol grubu” olarak 4 grup içeren bir çalışmada, kontrol grubunu 0, diğerlerini 1, 2, 3 ile, ya da kontrol grubunu 1, diğerlerini 2, 3, 4 ile kodlayabilirsiniz.

Bunun tersine kontrol grubu 4, diğer gruplar 1, 2, 3 diye de kodlanabilir. Aslında hangi gruba hangi rakamın verildiğinin istatistik olarak hiç önemi yoktur, çünkü nominal değişkenlerin düzeyleri arasında sıralama söz konusu değildir. Yukarıdaki kodlama şekli, istatistik analiz sırasında kolaylık sağlamak için önerilmiştir.

İkili nominal değişkenlerde eğer değişken düzeyleri “var, yok” ya da “evet, hayır” gibi ise, “var” ya da “evet”leri, yani olumlu yanıtları 1 ile, “yok” ya da “hayır”ları, yani olumsuz yanıtları 0 ile kodlamak uygunudur. Bazı istatistik analiz tekniklerinin bilgisayarda yapılabilmesi için değişkenlerin bu şekilde kodlanmış olması gereklidir.

Ordinal değişkenlerin düzeyleri arasında bir sıralama söz konusu olduğu için, bu tip değişkenleri kodlarken, değişkenlerin düzeylerine göre sıralama yapılmalıdır. Örneğin “uyku düzeniniz nasıldır?” sorusuna verilen yanıtların “çok kötü, kötü, fena değil, iyi, çok iyi” olarak beş farklı düzeyde olduğunu düşünelim. Bu değişken düzeylerinin 0, 1, 2, 3, 4 ya da 1, 2, 3, 4, 5 olarak kodlanması en uygunudur. Aslında istatistik analiz teknikleri açısından bakıldığında ikisi arasında fark olmadığı gibi, 0, 1, 2, 3, 4 yerine 3, 4, 7, 8, 12 ya da 4, 3, 2, 1, 0 şeklinde kodlanması da farklı değildir. Eğer sayısal değişkenlerden bazıları ya da tümü ordinal değişkenlere dönüştürülecekse, aynı kodlama düzenine uymak yararlıdır.

Yine pratik kolaylık açısından, bir veritabanındaki tüm nominal ve ordinal değişkenlerde aynı kodlama sistemine uyulması yararlıdır. Örneğin bazı değişkenlerin 0, 1, 2 ile bazılarının 1, 2, 3 ile kodlanması hatalı olmasa bile uygun değildir.

DEĞİŞKENLERİ ADLANDIRIRKEN DİKKAT EDİLECEK NOKTALAR

Genellikle, çalışma sırasında her deneğe ait veriler ayrı bir veri giriş formuna yazılarak toplanır ve çalışma bitiminde topluca veritabanına girilir. Veritabanını hazırlarken, önce her değişkene bir ad vermek gerekir. Değişken adlarının, değişken verilerinin neler olduğunu hatırlatıcı olması önemlidir. Özellikle çok sayıda değişken varsa, bu konu çok daha önemlidir.

Bilgisayar ortamındaki veritabanına girilecek değişkenlerin adlarını belirlerken bazı kısıtlayıcı kurallar vardır. Her yazılımdaki kısıtlamalar kendine özgüdür. Bu nedenle kullanacağınız veritabanı yazılımının el kitaplarına başvurmanız gereklidir.

VERİ GİRİŞİNDE HATA KONTROLÜ

Bir çalışmada verilerin elde edilmesinden bilgisayara girilip, istatistik analizinin yapılacağı ana kadar çeşitli aşamalarda hatalar olabilir. Çünkü hastane kayıtları, hasta dosyaları, laboratuvar kağıtları gibi dokümanlardan alınan rakamlar genellikle önce hazırlanmış formlara girilir. Daha sonra bu kağıtlardan bilgisayara aktarılır. Veri girişi sırasında hatalı yazma ya da yanlış tuşa basma gibi nedenlerle hatalı veri girişi yapılması olasılığı çok fazladır. Eğer veritabanı çok geniş değilse, ekrandan ya da kağıttan okuyup, formları karşılaştırarak kontrol etmek gerekir. Eğer fazla sayıda değişken ve/veya denek varsa, bu şekilde hata denetimi pratik değildir. Değişkenin özelliklerine göre yöntemler geliştirilebilir.

Istatistik analize başlamadan önce olası hataların fark edilmesi için sıklık tabloları ve histogramların gözden geçirilmesi yararlı olur (Verilerin grafik yöntemlerle değerlendirilmesi Bölüm 7'de anlatılmıştır). Örneğin hastanın yaş olarak 51 yerine 551 girilmişse, yaşa ait sıklık tablosunda ya da histogramda bunu fark etmek kolaydır. Ama 67 yerine 76 girilmişse, bunu fark etmek hemen imkansızdır. Bir diğer yöntem de, değişken değerlerinin en düşük ve en büyük değerlerinin gözden geçirilmesidir (Tüm istatistik yadınaların tanımlayııcı istatistik seçeneklerinde verilerin "minimum" ve "maksimum" değerlerini elde edebilirsiniz). Örneğin "hayır" için 0 ve "evet" için 1 değerini verdığınız bir nominal değişkenin en büyük değeri 2 olarak işaretleyip, hatalı veri girildiği anlaşılır. Sayısal değişkenler için bunu anlamak her zaman mümkün olmayabilir.

Üçüncü bir yöntem, hatalı veri girişinin denetiminde ideal değildir. Eğer veritabanı çok geniş değilse, aynı verileri iki kişinin farklı veritabanlarına girip, sonra bu iki veritabanının kullanılan yazılıma özgü yöntemlerle karşılaştırıl-

ması en ideal hata denetimi yöntemi olabilir.

Veri girişi hata denetimi kesinlikle aceleye getirilmemelidir. Aksi takdirde, istatistik analiz bittikten sonra farkedilecek bir hatalı veri girişi, tüm çabaların boş gitmesine neden olur.

Veritabanında hata denetimini yaptım; istatistik analize başlıyorum

Hayır! Hata denetimini yaptıktan sonra, derhal istatistik analize başlamak fazla aceleciğ olur. Çünkü değişkenlerin ortalama, ortanca, standart sapma vb. gibi temel tanımlayıcı istatistikleri görüldükten sonra, analitik istatistikte seçilecek yöntem değişimelidir. Verilerin dağılımına göre ya verilerde dönüşüm yapmak, ya da nonparametrik istatistik analiz yöntemlerini uygunlamak gerekebilir. Tanımlayıcı istatistik Bölüm 5'te, normal dağılım Bölüm 3'te anlatılmıştır.

SPSS'TE VERİ VE DOSYA İŞLEMLERİ

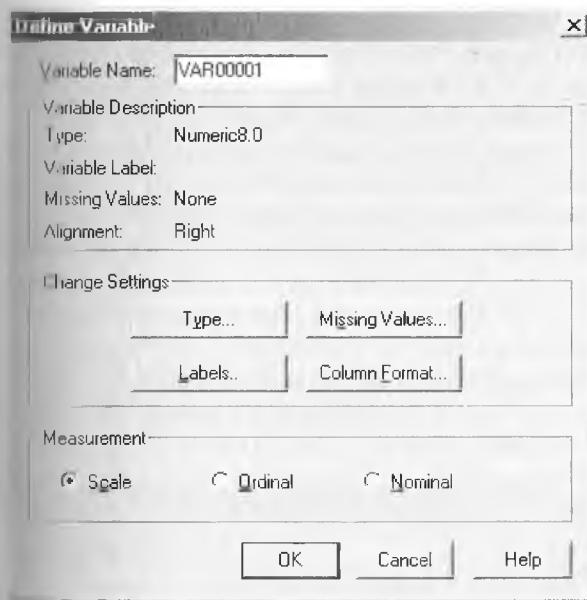
SPSS'i çalıştırığınızda veri girişi için *Data Editor* penceresi ve istatistik sonuçları, hata mesajları, vb. çıktıların yazılacağı *Viewer* penceresi açılır. Penceler arasında gidiş-gelişi ekranın en üstünde yer alan “*Window*” seçenekinden yapabilirsiniz.

VERİTABANI NASIL HAZIRLANIR?

SPSS veri penceresi, veri dosyası açmak ve düzenlemek (*edit*) için *spreadsheet* tarzında hazırlanmış bir penceredir. Ekranda görülen kısmı, soldan sağa ve yukarıdan aşağıya binlerce sütun ve satır olarak devam eden *spreadsheet*'in sol üst köşesidir. Bu penceredeki her satır, bir denek ya da gözlemi temsil eder. Örneğin bir hastaya ait tüm veriler tek satıra girilir. Her sütun bir değişkeni temsil eder. Örneğin hastaların yaşı, “yas” adı verilmiş olan değişkenin sütununa girilir.

Değişkenlerin tanımlanması

Değişkenlerin tanımlanması için “**Data >> Define variable**” ile **Define Variable** diyalog kutusu açılır. Değişkenin yerleştirileceği sütunun en üstündeki göze (var00001) çift klik yapıldığında da aynı pencere açılır (Şekil 1).



Şekil 1
SPSS'te "Define Variable" diyalog kutusu.

Değişken adı

Değişken adı *Variable name* yazan yere girilir. Değişkenlere ad verirken uyuşulması gereken bazı kurallar vardır:

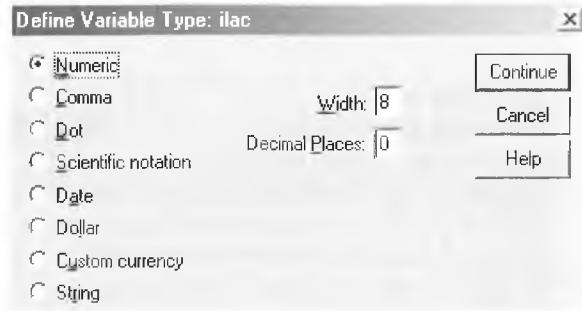
1. Değişken adları bir harfle başlamalıdır. Daha sonra harf, rakam, nokta ya da $_$, \$, # gibi bazı işaretler yer alabilir. "1SAATGLU", "\$UCRET" geçerli değişken adları değildir. "GLU1SAAT", "UCRET\$" geçerlidir.
2. Değişken adının son karakteri nokta olamaz. "SGOT." geçerli değildir. "SGOT.0" geçerlidir.
3. Değişken adı en fazla 8 karakter içerebilir.
4. Değişken adı içinde boşluk ya da ?, *, !, ' gibi işaretler yer alamaz.
5. Birden fazla değişkene aynı ad verilemez.
6. Değişken adındaki harflerin büyük ya da küçük harfle yazılmış olması farklı değildir. "LOKOSIT", "Lokosit" ve "LoKoSiT" aynı ad olarak işlem görür.
7. ALL, LT, AND, NE, BY, NOT, EQ, OR, GE, TO, GT, WITH, LE gibi sözcükler değişken adı olamaz. Bunun nedeni bu kısaltmaların SPSS programlama dilinin komutları olmasıdır.

SPSS'te değişken tanımlama diyalog kutusundan değişken adını belirlerken bu kurallara uyulmazsa, bir uyarı mesajı ile belirtilir.

Değişken tipi

Değişkenler için adından başka özellikler de eklenebilir. Bu özellikler (değişkenin tipi, etiketi, eksik değerlerin tanımı ve değişken değerlerinin gözdeki yerlesimi) *Variable Description* penceresinde görülür.

Change Settings penceresindeki *Type* seçeneği ile değişken tipinin belirleneceği **Define Variable Type** diyalog kutusu açılır (Şekil 2). Değişkenler önceden sayısal (*numeric*) olarak kabul edilmiştir. Sayısal değişkenler için geçerli değerler, başında artı ya da eksi işaretti de bulunabilen ve ondalık değer için nokta içeren rakamlardır. Noktanın sağında kaç rakam için yer ayrılacağı *decimal places*'te, değişken için ayrılan gözde ondalık noktası dahil kaç rakamın yer aldığı *width*'te belirlenir. Örneğin bir değişkenin alabileceği maksimum değer 9999.999 ise, "decimal places = 3" ve "width = 8" olarak girilmelidir. Sayısal değişkenlerde maksimum genişlik için 40, maksimum ondalık için 16 yer ayrılabilir. Ondalık için ayrılan yer yalnızca ekran görüntüsü içindir. Örneğin ondalık için 2 yer ayrılmış ve 13.456 girilmiş olsa, *spreadsheet*'te 13.46 görülecek, ama hesaplamalarda 13.456 kullanılacaktır.



Şekil 2

SPSS'te "Define Variable" diyalog kutusundan ulaşılan "Define Variable Type:" diyalog kutusu.

Noktalı sayısal (*dot*) ya da virgülü sayısal (*comma*) değişkenler daha çok muhasebe veri dosyalarında kullanılan bin, milyon, milyar vb.leri nokta ya da virgülerle ayırmak için kullanılır. Örneğin bir milyon, sayısal değişken seçilmişse 1000000, noktalı sayısal değişken seçilmişse 1.000.000, virgülü sayısal değişken seçilmişse 1,000,000 olarak görülecektir. Bu farklılık, yine yalnızca ekran görüntüsünde kolaylık sağlamak amacıyla hizmet eder. Hesaplamalarda her üç tip değişkendeki sayılar aynı şekilde işleme alınır.

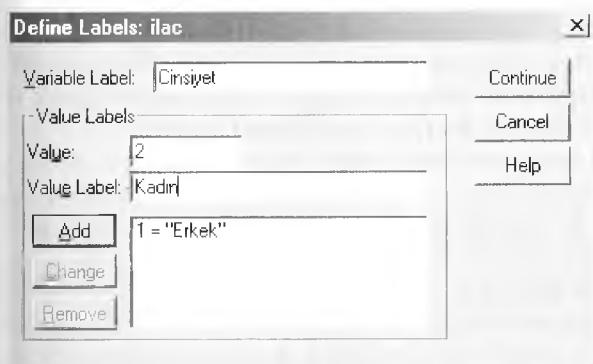
Karakter (*string*) değişkenlere ise her türlü sayı, harf, işaret girilebilir. Genişliği 8 ya da daha az olan karakter değişkenlere kısa karakter değişkenler, daha uzun değişkenlere uzun karakter değişkenler denir. Kısa karakter değişken-

ler bazı istatistik testlerde kullanılabilirken, uzun değişkenler kullanılmaz.

Diğer değişken tipleri tarih, dolar vb.dir. Tarih formatında tanımlanmış olan değişkenlere veri girişi yaparken gün ay ve yıl arasında /, \, boşluk, nokta, virgül konabilir. Örneğin değişken tipi "dd-mmm-yyyy" formatında bir tarih değişkeni ise (burada d, day = gün; m, month = ay ve y, year = yıl anlamındadır), örneğin 3 Aralık 1994 tarihini girmek için veri editörü penceresine "3.12.1994", "3/12/94", "3,12,94", "3 dec 1994" vb. çeşitli şekillerde girilen veri, kendiliğinden "03-DEC-1994"e dönüştürülür.

Değişken ve değer etiketleri

Change Settings'teki Labels seçeneği ile açılan **Define Labels** diyalog kutusundan değişken ve değer etiketleri eklenebilir ya da daha önceden eklenmişse, değiştirilebilir (Şekil 3).



Şekil 3
SPSS'te "Define Variable" diyalog kutusundan ulaşılan "Define Labels:" diyalog kutusu.

Bir değişkenin etiketli olup olmamasının, istatistik analiz açısından hiç bir önemi yoktur. Eğer değişken adları çok açıklayıcı değilse, etiket koymak yararlı olacaktır. Örneğin "Postoperatif üçüncü saatte periton sıvısında amilaz düzeyi" değişkeninin adı "PO3PRAMI" olarak konulmuşsa ve adı buna benzer başka değişkenler de varsa, "Çıktı" penceresinde karışıklığa yol açmak için bu değişkene "Postoperatif üçüncü saatte periton sıvısında amilaz düzeyi" etiketi eklenebilir. Değişken etiketi olarak maksimum 120 karakter olarak girilebilir. Değişken ve değer etiketlerinde her türlü işaret ve Türkçe harf yer alabilir.

Sayısal olmayan değişkenler de sıkılıkla sayısal değişkenler olarak tanımlanıp değerler kodlanır. Örneğin nominal bir değişken olan cinsiyeti "erkek" ve "kadın" yazmak yerine 1 ve 2 değerleri alan sayısal değişken olarak ta-

nımlamak daha kolaydır. Bu durumda 1'in erkek, 2'nin kadın olduğu etiketlenebilir. İstatistik işlemlerde değerler sayısal olarak alınır.

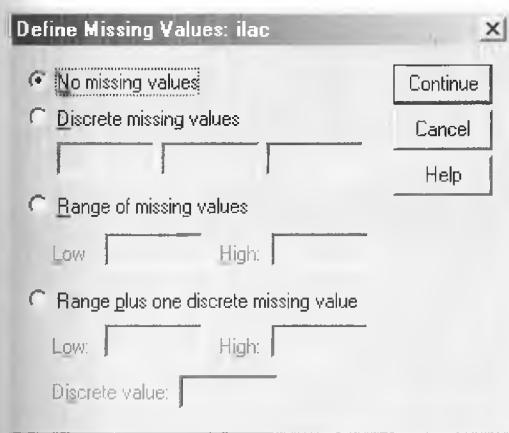
Değerleri etiketlendirmek için *Value labels* penceresindeki Value kısmına etiketlenecek değeri, *Value Label* kısmına ise değerin etiketi yazılır ve Add tuşuna basılarak, etiket listesine alınır (Şekil 3).

Kayıp değerler

Verileri toplarken verilerde kayıp olması çoğu zaman kaçınılmazdır. Bir hastanın kanının alındığı tüplerden biri kırılabilir ya da kaybolabilir; kobaylardan bir kısmı çalışma bitmeden ölebilir ya da deneklerden bir kısmı anketteki bazı sorulara yanıt vermemiş olabilir. Bu tip kayıp değer içeren deneklerin kayıp olmayan değerlerini kullanmak isterken, kayıp değerlerin de bir şekilde belirtilmesi gereklidir.

SPSS'te iki tip kayıp değer (*missing value*) tanımlanabilir. Veri girişi sırasında boş olarak geçen değerler "*System-missing value*" olarak algılanır ve ":" (nokta) ile gösterilir. Bunun dışında bazan belli değerler kayıp değer olarak belirlenebilir. Örneğin yaşı belli olmayan deneklerin yaşı 999 olarak giriliip, daha sonra "yaş" değişkeninin aldığı 999 değeri kayıp değer olarak tanımlanırsa, istatistik işlemlerinde bu denekler test dışında bırakılır.

Change Settings'ten ulaşılan **Define Missing Values** diyalog kutusunda No missing values önceden seçilmiş olan seçenekdir (Şekil 4). Bu durumda yalnızca "*system-missing value*" vardır. Bunun dışında ya bir, iki ya da üç farklı kayıp değer (örn. 99, 100 ve 101), ya kayıp değere karşılık gelecek değer aralığı (99 ila 120), ya da bir değer aralığı + bir değer (99 ila 120 ve ayrıca 150) olarak kayıp değerler tanımlanabilir. Sayısal ya da kısa karakter değişkenlerde tek değer kayıp olarak etiketlenebilir. Değer aralığı ise yalnız sayısal değişkenler için mümkündür.



Şekil 4
SPSS'te "Define Variable" diyalog kutusundan ulaşılan "Define Missing Values:" diyalog kutusu.

Sütun formatı

Değişken değerlerinin girildiği gözlerin genişliği ve değerlerin gözlerdeki yerleşim formatı *Change Settings*'teki *Column Format* seçeneğinden ulaşılan **Define Column Format** diyalog kutusundan belirlenebilir.

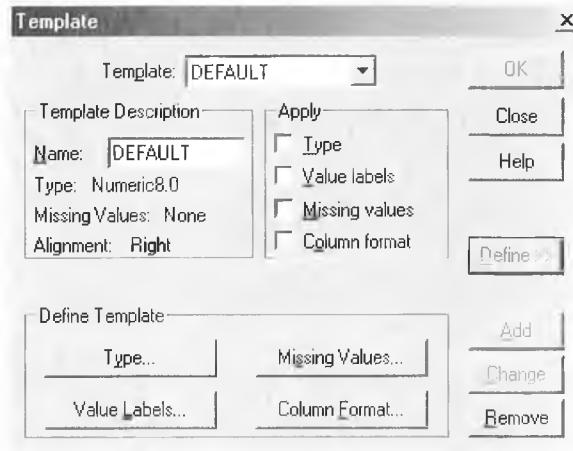
Ölçüm tipi

Değişkenin ölçüm tipi *Measurements*'taki *Scale*, *Ordinal* veya *Nominal*'den biri seçilerek belirlenebilir.

"TEMPLATE" NEDİR, NE İŞE YARAR?

Aynı şekilde tanımlanmış değişken özellikleri birden çok değişkene aktarılabilir. Örneğin 8 tane değişkenin özellikleri aynı ise (sayısal, 0, 1 ve 2 değerlerini alabilen, değer etiketleri 0=HAYIR, 1=EVET, 2=BAZEN, sütun formatı ve kayıp değer tanımı aynı) önce bu özelliklerde bir şablon (*template*) tanımlanır ve daha sonra bu şablonun özellikleri bu değişkenlere aktarılır.

Şablon tanımlamak için "Data >> Templates" seçeneği ile **Template** diyalog kutusu açılır (Şekil 5). *Define* ile şablon tanımlandıktan sonra *Change* ile şablon değiştirilir ya da yeni bir şablon adı yazıldıktan sonra *Add* ile yeni bir şablon tanımlanmış olur. *Apply* penceresinden şablonun hangi özelliklerinin seçilmiş değişkenlere aktarılacağı seçilir.



Şekil 5
SPSS'te "Template"
diyalog kutusu.

VERİ GİRİŞİ VE DÜZENLEMESİ (EDIT) NASIL YAPILIR?

Değişkenler yukarıdaki şekilde hazırlandığı zaman Veri penceresinde her sütunun en üstteki gözünde o sütunda hangi değişkenin yer aldığı yazılır. Değer etiketleri "1=kontrol ve 2=hasta" olan grup değişkeni, değer etiketleri "1=erkek ve 2=kadın" olan cinsiyet değişkeni ve yaşın girileceği yas değişkeninden ibaret bir veritabanı dosyası oluşturalım ve veri girişine başlayalım.

Birinci hasta kontrol grubundan 41 yaşında bir erkek olsun. Farenin normalde ok şeklinde olan imleci, *spreadsheet*'in üzerinde artı şekline dönecektir. Artı işaretini birinci satırın (yani birinci hastaya ait olan satırın) birinci sütunu (yani grup adlı değişkenine) getirin ve klikleyin. O gözün çerçevesinin çizgileri daha kalın hale gelecektir (Şekil 6).

The screenshot shows the SPSS Data Editor window titled "Untitled - SPSS for Windows Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Window, and Help. The toolbar contains various icons for data manipulation. A table is displayed with three columns: "grup", "cinsiyet", and "yas". Row 1 is highlighted in blue, indicating it is selected. The first column has values 1 through 5. The bottom status bar says "SPSS".

	grup	cinsiyet	yas
1			
2			
3			
4			
5			

Şekil 6
SPSS'te "Veri penceresi"nde veri girilecek gözün seçimi.

Birinci hasta kontrol grubundan olduğu için "1" yazın ve TAB ya da "sağa ok" tuşuna basın. Gözde "kontrol" yazılacak ve birinci satırın ikinci sütunu (yani birinci hastanın cinsiyetinin girileceği gözün) çerçevesinin çizgileri kalımlaşacaktır. Bu kez, artık farenin tuşuna basmaya gerek kalmadan veri girişi yapılacak göz aktif hale gelecektir (Şekil 7).

The screenshot shows the SPSS Data Editor window with the same table structure as Figure 6. The first row now contains the value "Kontrol" in the "cinsiyet" column. The other rows remain empty. The status bar at the bottom says "SPSS".

	grup	cinsiyet	yas
1	Kontrol		
2			
3			
4			
5			

Şekil 7
SPSS'te "Veri Penceresi"nde veri girişi.

TAB ve "sağa ok" ile bir sağdaki göze, Shift-TAB ve "sola ok" ile bir soldaki göze, vb. gibi tuşlarla *spreadsheet* üzerinde gezebilirsiniz.

CTRL, PgDn, PgUp, Home, End, vb. gibi tuşlar ve kombinelerini kullanımları, Windows altında çalışan diğer programlardaki gibidir. Birinci hastanın cinsiyetini "1" ve yaşını "41" olarak girdikten sonra, ekran görüntüsünün Şekil 8'deki gibi olması gereklidir.

	grup	cinsiyet	yas
1	Kontrol	Erkek	41
2			
3			
4			
5			

Şekil 8
SPSS'te "Veri penceresi"nde veri girişi.

HOME tuşuna basarsanız, birinci satırın ilk gözüne geri dönersiniz. ENTER tuşuna bastığınızda ise ikinci satırın birinci gözü veri girişi için hazır hale gelir. Değişken değer etiketlerinin ekranda görünmesini istiyorsanız, View seçeneklerinden *Value Labels*'ı aktif hale getirmeniz gereklidir. Bu durumda ekran görüntüsü Şekil 9'daki gibi olacaktır.

The screenshot shows the SPSS Data View window titled "Untitled - SPSS for Windows Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Window, and Help. Below the menu is a toolbar with various icons. The data table has three columns: "grup", "cinsiyet", and "yas". The first column contains numerical values (1-7). The second column contains categorical labels: "Kontrol", "Kontrol", "Hasta", "Kontrol", "Hasta", "Hasta", and "Hasta". The third column contains numerical values: 41, 23, 42, 59, 70, 53, and 56. The status bar at the bottom right shows "SF".

	grup	cinsiyet	yas
1	Kontrol	Erkek	41
2	Kontrol	Erkek	23
3	Hasta	Erkek	42
4	Kontrol	Kadın	59
5	Hasta	Kadın	70
6	Hasta	Erkek	53
7	Hasta	Kadın	56

Şekil 9

SPSS'te "Veri Penceresi"nin
"View >> Value Labels"
aktifken görünümü.

Eğer, değişken değer etiketlerinin yerine değerlere verdığınız kodların ekranda görünmesini istiyorsanız, View'den Value Labels seçeneğini kapatın (Şekil 10).

This screenshot is identical to Sekil 9, showing the same data table in the SPSS Data View window. The only difference is that the "Value Labels" option is disabled, resulting in the raw numerical codes (1-7) being displayed in the "cinsiyet" column instead of the categorical labels.

	grup	cinsiyet	yas
1	1	1	41
2	1	1	23
3	2	1	42
4	1	2	59
5	2	2	70
6	2	1	53
7	2	2	56

Şekil 10

SPSS'te "Veri Penceresi"nin "View >> Value Labels" aktif değilken görünümü.

TANIMLAYICI İSTATİSTİK

bölüm 5

Merkezi eğilim ölçütleri

Ortalama

Ortanca

Mod

Geometrik ortalam

Ne zaman hangi merkezi eğilim ölçütünü kullanalım?

Yayılma ölçütleri

Değer aralığı

Standart sapma ve varyans

Değişim (varyasyon) katsayısı

Persentil

Çeyreklerarası aralık

Ne zaman hangi yayılma ölçütünü kullanalım?

Standart hata

Standart sapma ile standart hata arasındaki ilişki

SPSS'te tanımlayıcı istatistik nasıl yapılır?

Frequencies menüsü

Statistics altmenüsü

Format altmenüsü

Descriptives menüsü

Altgrupların tanımlayıcı istatistikleri

Bir çalışmanın raporunu yazarken, yalnızca istatistik analiz sonuçlarını vermek yeterli olmaz. Örneğin antihipertansif bir ilacın etkinliğinin araştırıldığı bir çalışmada "A ilacı ile sistolik kan basıncı düşmüştür ($p=0.005$)" cümleinden, A ilacının etkisinin istatistiksel olarak önemli bulunduğu anlaşılmaktadır. Peki, sistolik kan basıncındaki düşme klinik olarak da önemli midir? Ortalama sistolik kan basıncı 163 mmHg'dan 120 mmHg'ya mı düşmüştür, yoksa 148 mmHg'ya mı düşmüştür? Her iki durumda da tedavi öncesi ve sonrası ortalama sistolik kan basıncı değerleri arasındaki farkın istatistiksel önemini belirten p değeri aynı olabilir, ama ortalama sistolik kan basıncı

nı 120 mmHg'ya düşüren bir ilaçın klinik olarak da etkili bir ilaç olduğu söylenebilirken, 148 mmHg'ya düşüren bir ilaçın etkinliği, istatistiksel olarak önemli bulunmuş olsa da, klinik olarak etkisini kuşkuyla karşılamak gereklidir.

Tedavi sonrasında sistolik kan basıncı değerinden başka, tedavi öncesindeki ortalama sistolik kan basıncını da bilmek yararlı, hatta gereklidir. Ortalama sistolik kan basıncı değeri tedavi ile 120 mmHg'ya düşmüştür, ama tedavi öncesindeki değerin 163 mmHg, ya da 210 mmHg olması, yine yorumu etkileyecektir. İlaç etkilidir, ama çalışma grubunun özellikleri bilinirse, "A ilacı, hafif hipertansiyonda etkilidir", ya da "ağır hipertansiyonda etkilidir" gibi daha doğru bir yorumu gidilebilir.

İşte, çalışma grubunun çeşitli özelliklerinin özetlendiği ortalama, oran, standart sapma vb. rakamlar topluluğuna tanımlayıcı istatistik adı verilir.

MERKEZİ EĞİLİM ÖLÇÜTLERİ

Değişken değerleri hakkında bilgi veren en önemli ölçütlerin başında merkezi eğilim ölçütleri (*measures of central tendency*) gelir. Tıpta kullanılan üç merkezi eğilim ölçüyü ortalama (*mean*), ortanca (*median*) ve mod'dur. Ortalama ve ortanca en sık kullanılan merkezi eğilim ölçütleridir.

Ortalama

Değerlerin toplamının denek sayısına bölünmesiyle elde edilen aritmetik ortalama en sık kullanılan merkezi eğilim ölçütüdür. X üzerinde çizgiyle (\bar{X}) gösterilir. Sayısal değişkenler için merkezi eğilim ölçüyü olarak ortalama kullanılır. Ordinal değişkenler için ortalama kullanılamaz. Örneğin 25 yeni doğanın doğum ağırlığının merkezi eğilim ölçüyü olarak ortalama kullanılabılır, ancak APGAR skorunun merkezi eğilim ölçüyü olarak ortalama kullanılamaz.

Ortalama daki önemli rakam sayısı, değişkenlerdeki önemli rakam sayılarından fazla olmamalıdır. Örneğin ortalama hemoglobin düzeyi 10.882 g/dL olarak değil, 10.9 g/dL olarak verilmelidir.

Ortalama, aşırı değerlerden etkilenir. Örneğin ortalama kolesterol değeri hesaplanacak olan 15 hastanın 14'ünün kolesterol düzeyi 120 mg/dL civarında, birinin kolesterol düzeyi 350 mg/dL ise, bu tek hastanın değeri nedeniyle ortalama yükselir.

Ortanca

Ortanca, küçükten büyüğe (ya da büyükten küçüğe) doğru sıralandığında tam ortadaki denegin degeridir. Denek sayısı çiftse, ortanca, ortada yer

alan iki deneğin değerlerinin ortalaması olarak hesaplanır. Örneğin yukarıdaki örnekte denek sayısı 15 değil de, 16 olsaydı, sekizinci ve dokuzuncu deneklerin kolesterol değerlerinin ortalaması, ortanca olacaktı.

ÖRNEK 21

Dokuz deneğin Hb değerleri 10-14 g/dL arasında yer almaktadır. Ortalama Hb değeri 12.0 g/dL, ortanca Hb değeri 12.4 g/dL'dir. Hb'i 12.0 g/dL olan bir denek yerinde 5.0 g/dL olan bir denek olsa, ortalama 12.0 g/dL'den 11.2 g/dL'ye düşer, ortanca ise değişmez. Çünkü ortanca, 9 denek sıralandığında ortadaki deneğin Hb değeridir. En düşük ya da en yüksek Hb değerleri ne olursa olsun, ortanca değeri değişmez.

Ortanca, ordinal veriler için en iyi merkezi dağılım ölçütüdür. Çünkü değerlerin sıralaması dikkate alınarak elde edilir. Sayısal veriler için tercih edilen merkezi eğilim ölçüttü ise ortalamadır. Simetrik dağılımlar için ortalama ve ortanca birbirine yakındır.

Ortanca aşırı değerlerden etkilenmez. Bu nedenle aşırı uç değerler varsa, sayısal veriler için de ortanca tercih edilmelidir.

Mod

Mod, değişken değerleri içinde en fazla görülen değerdir. Örneğin 100 kişinin Hb değerleri, sırasıyla oniki kişinin Hb'i 14.5 g/dL, on kişinin 14.0 g/dL, beş kişinin 13.5 g/dL, üç kişinin 15.5 g/dL ve kalan yetmiş kişinin Hb değerleri farklı değerler alıyorsa, bu grubun Hb değerlerinin modu 14.5 g/dL'dir. Mod, tipten nadiren kullanılan bir merkezi eğilim ölçütüdür.

Geometrik ortalama

Denek sayısı = n olduğunda, geometrik ortalama (*geometric mean: GM*), denek değerlerinin çarpımının $n^{\text{inci}} \text{kökü}$ olarak hesaplanır. Örneğin dört deneğin yaşları 24, 27, 38, 56 ise geometrik ortalama $\sqrt[4]{(24)(27)(38)(56)}$ 'dır. Hesaplama kolaylığı açısından aşağıdaki formül daha uygundur.

$$\log(\text{GM}) = \sum \frac{\log X}{n}$$

Geometrik ortalama, asıl olarak logaritmik dağılım gösteren veriler için kullanılır. Aşırı değerlerin olduğu sayısal verilerin merkezi eğilim ölçüttü olarak da kullanılabilir.

Ne zaman hangi merkezi eğilim ölçütünü kullanalım?

Bir değişkenin merkezi eğilimini en iyi temsil eden ölçütü seçerken, göz önüne alınması gereken iki önemli etken vardır. Bunlar değişkenin ölçüm skaları (yani değişkenin ordinal mı, sayısal mı olduğu) ve değişken değerlerinin dağılımıdır. Değişken değerlerinin dağılımı simetrik ya da çarpık (*skewed*) olabilir. Eğer aşırı düşük ya da yüksek değerler varsa, dağılım simetrik değildir. Uygun merkezi eğilim ölçütlerinin seçimi şu şekilde yapılabilir:

- Simetrik dağılan sayısal veriler için ortalama,
- Ordinal veriler ya da simetrik olmayan sayısal veriler için ortanca,
- Logaritmik skalada ölçülen veriler için geometrik ortalama seçilmelidir.

YAYILMA ÖLÇÜTLERİ

Farklı grupların ortalama ve ortanca gibi merkezi dağılım ölçütleri aynı olduğu halde, gruplar birbirlerinden çok farklı olabilirler. Bu nedenle merkezi eğilim ölçütleri yanında, yayılma ölçütleri (*measures of dispersion*) de çok önemlidir. Çeşitli yayılma ölçütleri vardır.

ÖRNEK 22

Her birinde yedi hasta bulunan iki gruptaki serum kolesterol değerleri birinci grupta 150, 160, 160, 160, 160, 170 ve 170 mg/dL, ikinci grupta ise 80, 100, 140, 160, 180, 225 ve 245 mg/dL'dir. Her iki grubun da ortalama kolesterol değeri 161 mg/dL ve ortanca kolesterol değeri 160 mg/dL'dir. Oysa iki grupta yer alan hastalar, serum kolesterol düzeyleri açısından hiç de benzer değildir.

Değer aralığı

Değişken değerlerinin dağılımını belirten yararlı bir ölçüt, en büyük ve en küçük değer arasındaki fark olan değer aralığıdır (*range*). Yukarıdaki örnekte değer aralığı, birinci grup için 20, ikinci grup için 165'tir. Değer aralığı da, ortalama gibi üç değerlerden çok etkilenir. Üstelik en uçtaki iki değer arasında kalan değerler hakkında bilgi vermez.

Standart sapma ve varyans

Tüm değerlerin dağılımı ile ilgili bilgi veren, sık kullanılan yayılma ölçütleri standart sapma (*standard deviation*) ve varyanstır (*variance*). Tüm değerler eşitse, her ikisi de sıfıra eşittir. Değerler arasında farklar arttıkça standart sapma ve varyans büyür.

SD, sd ya da s ile gösterilen standart sapma, değişken değerlerinin ortalamanın etrafındaki yayılmasını temsil eden bir yayılma ölçütüdür. Formülü şöyledir:

$$SD = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}}$$

\bar{X} = Ortalama

n = Denek sayısı

Standart sapmanın karesine varyans adı verilir. Standart sapma ve varyans, birçok istatistik analiz yönteminin temelini oluşturur.

Dağılım özelliği ne olursa olsun, değerlerin en az %75'i ortalama ± 2 SD içinde yer alır.

ÖRNEK 23

Yüz kişinin ortalama Hb değeri 12.3 g/dL, ve SD'sı 0.3 g/dL olarak hesaplanmıştır. Bu gruptaki deneklerden en az 75'inin Hb değeri 11.7 ile 12.9 g/dL arasında yer almmalıdır.

Normal dağılım gösteren değişken değerleri için aşağıdaki kurallar geçerlidir:

- Değerlerin %67'si ortalama ± 1 SD içindedir.
- Değerlerin %95'i ortalama ± 2 SD içindedir.
- Değerlerin %99.7'si ortalama ± 3 SD içindedir.

Değişim (varyasyon) katsayısı

Tablo 1'de görüldüğü gibi, normal servise yatan hastalarla karşılaştırıldığında, yoğun bakım servislerinde yatanların ortalama Hb değeri daha düşük olarak gözlenirken, kolesterol değerleri arasında belirgin bir fark yok. Ortalamların birbirinden farklı olup olmadığı konusunu bırakıp, değişken değerlerinin yayılma özelliklerine bakalım. Hb'e baktığımızda, yoğun bakımda yatanların SD'sı, normal serviste yatan hastaların Hb değeri SD'ndan daha büyük. O halde yoğun bakımda yatan hastalardaki Hb değerlerinin içinde, ortalama Hb değerinden uzakta bulunan (daha düşük ya da yüksek) değerler

Bir hastanede normal servislere ve yoğun bakım servislerine yatan hastalarda klinik ve laboratuvar ölçümler yapılmıyor. Farklı servislere yatan hastalarda bu ölçümlerin farklılık gösterip göstermediği araştırılıyor. Basitleştirmek için yalnızca Hb ve kolesterol değerlerini ele alalım. Her iki gruptaki hastaların Hb ve kolesterol değerlerinin ortalama ve standart sapmaları Tablo 1'de verilmiştir.

bulunmaktadır. Dağılımin normal olduğunu varsayırsak, normal servis hastalarının %95'inin Hb değeri 9.9-14.7 g/dL arasında iken, yoğun bakım hastalarının %95'inin Hb değeri 6.8-15.2 g/dL arasındadır. Yoğun bakım hastalarının Hb değerleri daha geniş bir aralığa yayılmıştır. Kolesterol değerleri için de aynı yorumu yapabiliriz. Bu sonuca ulaşmak için SD'ları karşılaştırmak yeterli oldu. Çünkü her iki grubun aynı birimle ölçülen değişkenlerini karşılaştırdık.

Tablo 1 Örnek 24'teki çalışmanın sonuçları.

	Normal Servis		Yoğun Bakım	
	Ortalama	SD	Ortalama	SD
Hb (g/dL)	12.3	1.2	11.0	2.1
Kolesterol (mg/dL)	134	45	137	67

Şimdi yalnız normal servis grubunu alalım ve Hb ve kolesterol değerlerinin yayılma derecesini karşılaştıralım. SD'sı 45 olduğuna göre, kolesterol değerinin Hb değerlerine göre çok daha geniş bir aralığa yayıldığını söyleyebilir miyiz? Elbette hayır. Çünkü Hb ve kolesterol değerlerinin ölçüm birimleri farklıdır.

Birimleri farklı olan değişkenlerin yayılmalarını karşılaştırmak için değişim katsayıları kullanabilir. **Değişim katsayısı (coefficient of variation)**, standart sapmanın ortalamaya oranının yüzde olarak ifadesidir. **Kısaca cv ile gösterilir**. Örneğin normal servis grubunun Hb değerlerinin değişim katsayısı $(1.2/12.3) \times 100 = \%9.8$ 'dir. Kolesterol değerlerinin varyasyon katsayısı ise $(45/134) \times 100 = \%33.6$ 'dır. İşte şimdi kolesterol değerlerinin Hb değerlerine göre daha geniş bir aralığa yayıldığını söyleyebiliriz.

Değişim katsayısı, biyolojik bilimlerde, genellikle bir laboratuvar işleminin ölçüm doğruluğunu test etmek için kullanılır.

ÖRNEK 25

Serumda kalsiyum düzeyi tayini için geliştirilen yeni bir yöntemle hazırlanmış bir kitin güvenilirliği belirlenmek isteniyor. Aynı kişiden alınan serum örneğinde bu kit kullanılarak 20 kez kalsiyum ölçümü yapılıyor ve 20 ölçüm sonucundaki değerlerin ortalaması 10.5 mg/dL ve SD'sı 0.5 mg/dL olarak bulunuyor.

Örnek 25'teki verilerle hesaplanan cv %4.7'dir. Genellikle laboratuvar yöntemlerin kabul edilebilir maksimum cv'sı %10'dur. Buna dayanarak bu kalsiyum kiti ile güvenilir ve tekrarlanabilir ölçümler yapılabildiği söylenebilir. Aynı serum örneğinden yapılan ölçümlerin %95'inin $10.5 \pm 1.0 = 9.5$ ila 11.5 arasında olması beklenir. Eğer SD, 0.5 değil de, örneğin 1.05 olsaydı (yani cv %10 olsaydı), ölçümlerin %95'i $10.5 \pm 2.1 = 8.4$ ila 12.6 arasında olabilirdi ki, bu durumda sonuçlar yeterince güvenilir olarak kabul edilemezdi.

Persentil

Yüz kişinin boy ortalamasının 168 cm olduğunu ve bunların içinde en kısa olanının 152 cm, en uzun olanının 191 cm olduğunu varsayıyalım. Bu kişileri kısaltan uzuna doğru boy sırasına dizdiğimizde, 95'inci kişinin boyu, bu grubun boy değerlerinin 95'inci persentiline (percentile) eşittir. Aynı şekilde 5'inci kişinin boyu da 5'inci persentile eşittir. Ellinci persentil tam ortadaki kişinin değeridir, yani ortancaya eşittir.

Persentil genellikle, boy, ağırlık vb. standart normların dağılımları için kullanılır. Bir diğer kullanıldığı yer ise, laboratuvar değerlerinin alt ve üst normal sınırlarının belirlenmesidir. Genellikle kabul edilen alt normal sınır 2.5 persentil ve üst normal sınır 97.5 persentildir.

Değerlerin dağılımı normalse, ortalama $-2SD$ ve ortalama $+2SD$, sırasıyla 2.5 ve 97.5 persentil değerlerine karşılık gelir.

ÖRNEK 26

Laboratuvara kullanılan serum demir düzeyi tayini yöntemi değiştirilmiştir ve yeni yöntemin normal alt ve üst sınırları bilinmemektedir. Bu sınırları belirlemek amacıyla 200 sağlıklı kişiye serum demir düzeyi ölçülüyor. En düşükten en yükseğe doğru sıraya koyulduğunda 5'inci ve 195'inci kişilerin serum demir düzeyi değerleri yeni yöntemin alt ve üst sınırları olarak belirleniyor.

Örnek 26'daki yaklaşımıla belirlenen alt ve üst sınırlara göre bundan böyle serum demiri ölçülecek sağlıklı deneklerin %95'i normal sınırlar içinde yer alacak, yalnız %5'inde ise denekler sağlıklı olduğu halde serum demir düzeyi

yinin normalden düşük ya da yüksek olduğu rapor edilecektir.

Çeyreklerarası aralık

25. ve 75. persentil değerleri arasındaki farka çeyreklerarası aralık (*inter-quartile range*) adı verilir. Yani değerlerin ortada yer alan %50'si, çeyrekler arası aralıktır.

Ne zaman hangi yayılma ölçütünü kullanalım?

Merkezi dağılım ölçütlerinin seçiminde olduğu gibi yayılma ölçütlerinin seçiminde de göz önüne alınması gereken bazı noktalar vardır. Aşağıdaki kurallar genellikle geçerlidir.

- **Merkezi eğilim ölçüyü** olarak **ortalama** kullanıldığı zaman, **yayılma ölçüyü** olarak da **standart sapma** kullanılır.
- **Persentil ve çeyrekler arası aralık** iki durumda kullanılır: (a) **Merkezi eğilim ölçüyü** olarak ortanca kullanılmışsa; (b) **Merkezi eğilim ölçüyü** olarak **ortalama** kullanılmış, ama gözlemlerin standart normlara uygunluğu karşılaştırılmak isteniyorsa.
- Dağılımin şekli ne olursa olsun değerlerin ortada kalan %50'si tanımlanmak isteniyorsa **çeyrekler arası aralık** kullanılır.
- Sayısal verilerde aşırı uç değerler belirtilmek isteniyorsa **değer aralığı** kullanılır.
- Sayısal verilerde farklı birimlerle ölçülen değerlerin dağılımları karşılaştırılmak isteniyorsa, **değişim katsayısı** kullanılır.

STANDART HATA

ÖRNEK 27

Koroner arter tıkanıklığı olan hastalarda kolesterolden fakir diyetin arter çapına etkisini ortaya koymak amacıyla planlanan bir çalışmada 50 koroner arter hastasına kolesterolden fakir diyet başlanmadan önce ve iki yıl sonra koroner arter çaplarının ölçülüyor ve arter çapındaki artma miktarı hesaplanıyor.

Kolesterolden fakir diyetin koroner arter tıkanıklığı olan hastalarda arter çapına etkisinin araştırılmasının amaçlandığı bu çalışmanın hedef evreni, koroner arterlerinde tıkanıklık olan hastalardır. Bu evrenden seçilecek bir örneklem, bir hastaneye başvuran koroner arter hastalarından oluşan 50 kişilik bir grup olabilir. Bir örneklemdeki denek değerlerinin örneklem ortalamasını

dan aşağı ya da yukarı olmak üzere ne kadar uzaklaştığının, yani denek değerlerinin yayılmasının ölçütünün standart sapma olduğunu biliyorsunuz. Bu çalışmada arter çapındaki artma miktarının ortalamasının %15 ve standart sapmasının %3.5 olarak bulunduğu düşünürsek, hastadan hastaşa olan değişkenlik hakkında bir fikir edinmiş oluruz. Koroner arter hastaları evreninden aynı büyülükte, yani 50 kişilik başka bir örneklem seçildiğinde, arter çapındaki artma miktarı %15 değil de, %14 ya da %16 olabilir. Yani bir örneklemdeki deneklerin değişkenliği yanında, farklı örneklemeler arasında da değişkenlik olabilir. **Aynı evrenden seçilecek, ya da seçilmesi mümkün olan aynı büyülükteki örneklemelerin ortalamalarının yayılmasına ortalamanın örneklem dağılımı (sampling distribution of mean) adı verilir.**

Ortalamanın örneklem dağılıminin ölçütü ortalamanın standart hatasıdır (standard error of mean). Kisaca **SEM** ile gösterilen standart hata, standart sapmanın denek sayısının kareköküne bölümüne eşittir.

$$\text{SEM} = \text{SD} / \sqrt{n}$$

SD = Standart sapma
n = Denek sayısı

Yukarıdaki örnek için SEM, $3.5 / \sqrt{50} = 0.05$ 'dir.

Yukarıdaki örnekte ortalama %15 ve standart sapma %3.5 olduğuna göre, hastaların %95'inde arter çapındaki artma miktarının %8 ile %22 ($15 \pm 2\text{SD}$) arasında olduğu kestirilebilir. Benzer şekilde SEM %0.5 olduğuna göre, evrenden seçilecek 50 kişilik örneklemelerin %95'inin ortalamaları %14 ile %16 ($15 \pm 2\text{SEM}$) arasında olacaktır.

Standart sapma ile standart hata arasındaki ilişki

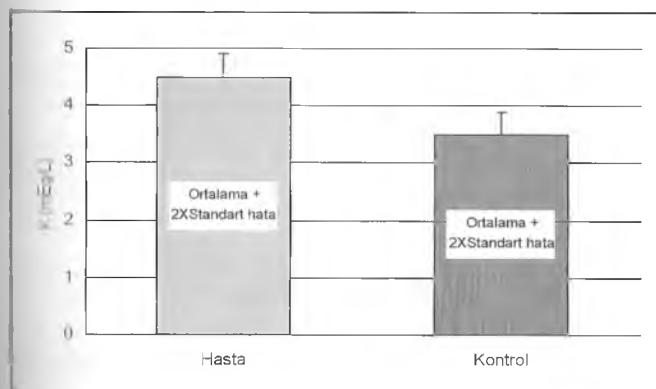
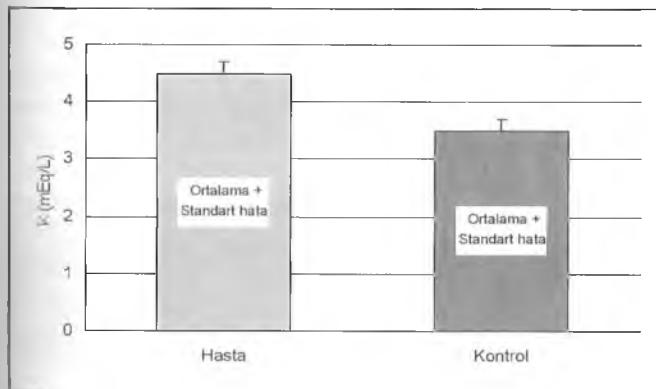
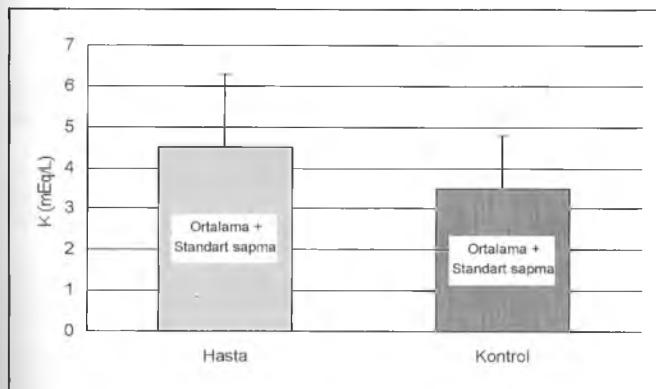
Standart sapmanın örneklem değerlerinin ortalamadan sağa ve sola (ya da aşağı ve yukarı) ne kadar dağıldığını gösteren bir ölçüt olduğunu biliyorsunuz. Dağılım eğrisi ne kadar basıksa (yani platikurtikse), standart sapma o kadar fazla, dağılım eğrisi ne kadar sivriyse (yani leptokurtikse) standart sapma o kadar azdır. Normal dağılımda tüm değerlerin yaklaşık %68'i maksimum bir standart sapma, %95'i ise iki standart sapma kadar ortalamadan uzaktır. Örneğin 10000 kişinin K değeri ortalamasının 4.2 mEq/L ve standart sapmasının 0.5 mEq/L olduğunu varsayıyalım. O halde 6800 kişinin K değeri 3.7-4.7 mEq/L arasındadır. 9500 kişinin K değeri ise 3.2-5.2 mEq/L arasındadır. Kalan 500 kişinin yarısının K değeri 3.2 mEq/L'den küçük, yarısının ise 5.2 mEq/L'den büyültür. Bu kez K değeri ortalaması aynı, ama standart sap-

masının 0.1 mEq/L olursa, 9500 kişinin K değerlerinin 4.0-4.4 mEq/L arasında yer almaktan olduğu, yani standart sapma daha küçük olduğu için, değerlerin ortalaması K değeri civarında toplandığı anlaşılır.

Standart hata (standart sapmanın denek sayısının kareköküne bölümlüne eşit olması nedeniyle), standart sapma arttıkça artar, denek sayısı arttıkça azalır. Standart hata, örneklem ortalamasının, aynı evrenden alınması mümkün olan eşit büyülükteki örneklemelerin ortalamalarının değişkenliği hakkında fikir verir. Örneğin 10000 kişi içinden 100 kişi alıp, K değerine bakıldığını ve ortalaması 4.1 mEq/L, standart sapma 0.5 mEq/L olduğunu varsayıyalım. Bu durumda standart hata $0.5 / \sqrt{10} = 0.05$ mEq/L olacaktır. Bu değer 10000 kişilik evrenden alınabilecek tüm 100 kişilik örneklemelerin K ortalamasının dağılıminin bir ölçütüdür. 100 kişilik örneklem ortalamalarının dağılım eğrisi çizildiğinde, ortalamaların %95'inin bu örneklem ortalamasından maksimum iki standart hata uzaklıkta yani 4.0-4.2 mEq/L arasında olması beklenir.

Makalelerde ortalaması ile birlikte bazen standart sapma, bazen de standart hatanın verildiği görülmektedir. **Çalışmaya alınan örneklemin yayılma özellikleri verilmek isteniyorsa, doğru olanı standart sapmanın verilmesidir.** Çünkü standart hata, örneklemdeki deneklerin yayılımı ile ilgili olmadığı için, makaledeki çalışma grubunun değişkenliğini göstermez. Örneğin, multipl myelomlu 81 hastanın yaş ortalamasının 50 ve standart sapmanın 9 olarak bulunduğu bir çalışmada bu sonucun "ortalama \pm SEM" olarak yani 50 ± 1 olarak verildiğini düşünün. **Makaleyi okuyan kişinin istatistik kavramları hakkındaki bilgisi yeterli değilse, hastaların büyük çoğunluğunun yaşıının 50'ye çok yakın olduğunu düşününebilir. Oysa hesaplamırsa, hastaların %95'inin yaşı 32 ile 68 arasındadır.**

Çalışma gruplarındaki **ortalamaların karşılaştırıldığı** grafiklerde ise **standart hatanın kullanılması** daha doğrudur. Çünkü ortalamaların karşılaştırılması sırasında farklı grupların (ya da örneklemelerin) aynı evrenden seçilmiş seçilmediği test edilir. Bir grubun seçildiği evrenin ortalaması ile diğer grubun seçildiği evrenin ortalaması farklı ise grup ortalamalarının da farklı olduğu sonucuna varılır. Örneğin hasta grubunun K değeri 4.5 mEq/L, standart hatası 0.2 mEq/L, kontrol grubunun K değeri 3.5 mEq/L, standart hatası 0.2 mEq/L olsun. Gruplar tamamen normal dağılım gösteriyorlarsa, deneklerin %95'inin K değeri hasta grubunda 4.1-4.9 mEq/L, kontrol grubunda ise 3.1-3.9 mEq/L arasında olacaktır. Bu iki aralık kesişmediği için grupların K değerlerinin ortalaması istatistiksel olarak da farklı olacaktır (tabii değerlerin dağılımı tamamen teorik normal dağılıma uyuyorlarsa). Bu nedenle grafiklerde grup ortalamaları yanında ortalama $+2\text{SEM}$ ve ortalama -2SEM değerlerinin verilmesi doğrudur (Şekil 1).



Şekil 1
Çubuk yükseklikleri grup ortalamalarını göstermektedir. Üstteki şekilde standart sapma, ortadaki şekilde standart hata, alttaki şekilde standart hatanın iki katı yüksekliğinde hata çubuğu (*error bar*) yerleştirilmiştir. Doğru gösterim şekli alttaki şekilde gibidir.

SPSS'TE TANIMLAYICI İSTATİSTİK NASIL YAPILIR?

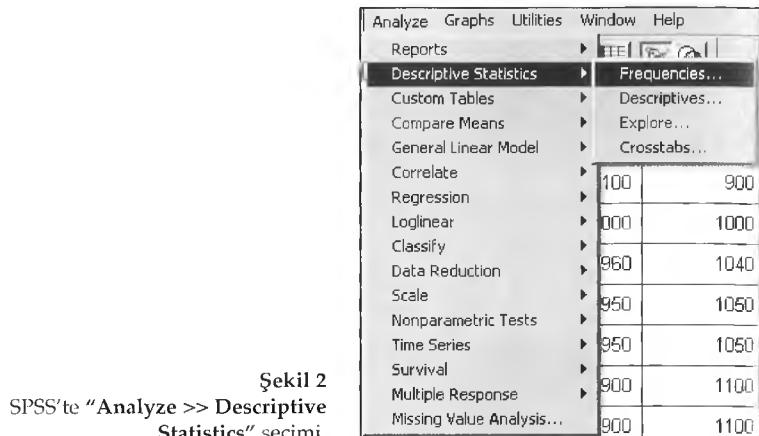
SPSS'de tanımlayıcı istatistik, Statistics menülerinin hemen hemen hepsi bulunur. Ama asıl olarak “Analyze >> Descriptive Statistics” seçenekindeki Frequencies ve Descriptives menüleri yardımcı ile yapılır.

Frequencies menüsü

ÖRNEK 28

Bir çalışmada kronik aktif hepatit (KAH), siroz, hepatosellüler kanser (HCC), bilier siroz ve karaciğer dışı hastalıklara bağlı asiti olan kişiler olmak üzere toplam 48 hastada asit sıvısında kalsiyum (Ca), demir (Fe), bakır (Cu), çinko (Zn) ve magnezyum (Mg) düzeylerine bakılıyor.

Bu çalışmada ilk yapılacak iş, çalışmaya alınan hastaların tanımlayıcı istatistiği olmalıdır. Bunun için Frequencies seçeneği seçilmelidir (Şekil 2).



Şekil 2
SPSS'te “Analyze >> Descriptive Statistics” seçimi.

Analyze >> Descriptive Statistics

Yukarıdaki seçimle Frequencies diyalog kutusu açılır. Bu kutudaki kaynak değişken listesinde sayısal ve kısa karakter değişkenler bulunur. *Display frequency tables* seçeneği önceden seçilmiştir. Bir değişkeni seçip OK'larseniz, o değişkenin sıklık tablosu elde edilir (Şekil 3).

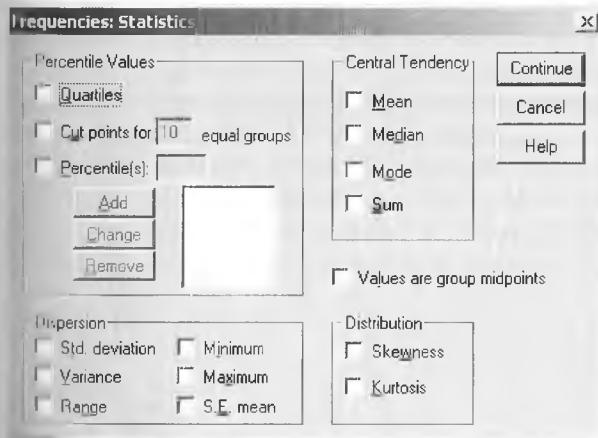
GRUP					
	Frequency	Percent	Valid Percent	Cumulative Percent	
Valid	1 Normal	14	29.2	30.4	30.4
	2 KAH	11	22.9	23.9	54.3
	3 Siroz	12	25.0	26.1	80.4
	4 HCC	5	10.4	10.9	91.3
	5 Bilier	4	8.3	8.7	100.0
	Total	46	95.8	100.0	
Missing	System	2	4.2		
	Total	48	100.0		

Şekil 3
Örnek 28'deki çalışmada hastaların gruplara göre dağılımının sıklık tablosu.

Şekil 3'te grup adlı sayısal değişkenin sıklık tablosu görülmektedir. En sol tarafta sütunda değer etiketleri görülmektedir. Frequency sütununda her gruptaki hasta sayıları yer almaktadır. Missing satırında iki hasta bulunduğu görülmektedir. Yani iki hastanın hangi gruba girdiği belirtilmemiştir. Percent sütununda bu iki hasta da dahil olmak üzere hesaplanan oranlar yüzde olarak verilmiştir. Valid percent sütununda ise bu iki hasta çıkarıldıktan sonra hesaplanan oranlar verilmiştir. Cumulative percent sütununda ise Valid percent sütundaki oranlara dayanılarak hesaplanan kümülatif oranlar vardır.

Statistics altmenüsü

Sayısal değişkenlerle ilgili tanımlayıcı istatistik için Frequencies diyalog kutusundan Statistics'i seçin. Böylece Frequencies: Statistics diyalog kutusu açılır. Bu kutuda dört ayrı grupta tanımlayıcı istatistik seçenekleri vardır (Şekil 4).



Şekil 4
SPSS'te "Frequencies: Statistics" diyalog kutusu.

1. Percentile values (Percentil değerler)

Quartiles (Çeyrekler). 25, 50 ve 75'inci persentil değerlerini verir.

Cut points for ... groups. Burada önceden 10 grup seçilmişdir. Yani 10 eşit gruba ayırarak, 10, 20, 30, ..., 90'inci persentilleri verir. "10" rakamını 2 ila 100'e değiştirerek, istediğiniz eşit aralıklarla persentil değerlerini alabilirsiniz.

Percentiles (Persentiller). Buraya istediğiniz persentil değerlerini girebilirsiniz. Her girdığınız değerden sonra *Add'i* seçip, yazdığınız değeri listeye ekleyin. Değerler listede küçükten büyüğe sıralı olarak dizilir. Listeye yanlış girdığınız değeri düzeltmek için, önce o değeri seçin, sonra düzeltin ve *Change'i* seçin. Listeden çıkarmak istediğiniz bir değer varsa, seçip, *Remove* edin.

2. Dispersion (Yayılma)

Bu kutuda standart sapma, varyans, değer aralığı, en küçük değer, en büyük değer ve ortalamanın standart hatası seçenekleri vardır.

3. Central tendency (Merkezi eğilim ölçütleri)

Bu kutuda ortalama, ortanca, mod ve değerler toplamı seçenekleri vardır.

4. Distribution (Dağılım)

Bu kutuda *skewness* (çarpıklık) ve *kurtosis* (basıklık) seçenekleri vardır.

Şekil 5'te tüm deneklerde Ca değişkenine ait değerlerin tanımlayıcı istatistikine ait sonuçlar vardır. Çeyrekler dışında 5'inci ve 95'inci persentiller de seçilmiştir.

Format altmenüüsü

Bu seçenek yalnız bu diyalog kutusunda değil, daha bir çok kutuda da vardır. Hepsinde de aynı işe yarar. Sıklık tablosunun ve sayfanın düzenini bu seçenekten ayarlayabilirsiniz.

Bu seçeneğin seçilebilmesi için *Display frequency tables* seçeneğinin seçilmiş olması gereklidir. Sıklık tablosundaki değerlerin nasıl sıralanacağını *Order by* kutusundan ayarlayabilirsiniz. Bu kutudaki dört seçenekten yalnız biri seçilebilir. Değişken değerlerinin küçükten büyüğe sıralandığı *Ascending values* önceden seçilmiştir. *Descending values* ile büyükten küçüğe, *Ascending counts* ile kategorilere düşen hasta sayısına göre azdan çoga, *Descending counts* ile çoktan aza doğru sıralanır.

Statistics		
KALSIYUM		
N	Valid	48
	Missing	0
Mean		3.720
Std. Error of Mean		.120
Median		4.000
Mode		3.0
Std. Deviation		.829
Variance		.687
Skewness		-.384
Std. Error of Skewness		.343
Kurtosis		-1.007
Std. Error of Kurtosis		.674
Range		2.7
Minimum		2.0
Maximum		4.7
Sum		178.6
Percentiles	5	2.000
	25	3.000
	50	4.000
	75	4.646
	95	4.693

Şekil 5
Örnek 28'deki çalışmada normal gruptaki Ca değerlerine ait tanımlayıcı istatistikler.

Şekil 3'te *Ascending values*'e göre gruplar sıralanmıştır. Şekil 6'da ise sırasıyla *Descending values*, *Ascending counts* ve *Descending counts*'a göre sıralanmış üç sıkılık tablosu görülmektedir. İlk tabloda grupların kodları 5'ten 1'e doğru sıralanmıştır. İkinci tabloda gruppardaki hasta sayılarına göre azdan çoka doğru ve son tabloda ise çoktan aza doğru sıralanmıştır.

GRUP

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	5 Bilier	4	8.3	8.7	8.7
	4 HCC	5	10.4	10.9	19.6
	3 Siroz	12	25.0	26.1	45.7
	2 KAH	11	22.9	23.9	69.6
	1 Normal	14	29.2	30.4	100.0
	Total	46	95.8	100.0	
Missing	System	2	4.2		
	Total	48	100.0		

GRUP

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	5 Bilier	4	8.3	8.7	8.7
	4 HCC	5	10.4	10.9	19.6
	2 KAH	11	22.9	23.9	43.5
	3 Siroz	12	25.0	26.1	69.6
	1 Normal	14	29.2	30.4	100.0
	Total	46	95.8	100.0	
Missing	System	2	4.2		
	Total	48	100.0		

Şekil 6
Örnek
28'deki
çalışmadaki
hastaların
gruplara göre
dağılımının
“Descending
values”,
“Ascending
counts” ve
“Descending
counts” a
göre sıralan-
mış sıklık
tabloları.

GRUP

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Normal	14	29.2	30.4	30.4
	3 Siroz	12	25.0	26.1	56.5
	2 KAH	11	22.9	23.9	80.4
	4 HCC	5	10.4	10.9	91.3
	5 Bilier	4	8.3	8.7	100.0
	Total	46	95.8	100.0	
Missing	System	2	4.2		
	Total	48	100.0		

Descriptives menüsü

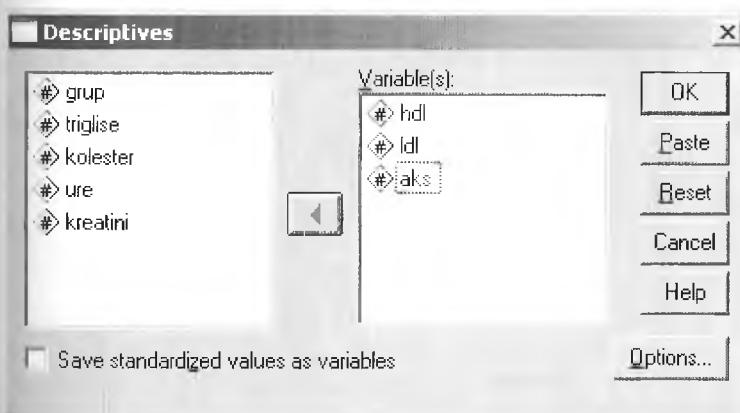
ÖRNEK 29

Tip I diabetes mellitus (DM), Tip II DM ve normal deneklerde kanda trigliserid, kolesterol, HDL, LDL, kan şekeri, üre ve kreatininin ölçüldüğü ve gruplar arasında fark olup olmadığına araştırıldığı bir çalışma yapılmıştır. Bu ölçümlerin tanımlayıcı istatistikleri belirtilmek isteniyor.

Gruplardaki değerlerin karşılaştırıldığı analitik istatistiğe geçmeden önce tanımlayıcı istatistikleri elde etmek gereklidir.

Analyze >> Descriptive Statistics

Yukarıdaki seçimle Descriptives diyalog kutusu açılır (Şekil 7). Bu kutuda kaynak değişken listesinde sayısal değişkenler bulunur. Tanımlayıcı istatistiklerini istediğiniz en az bir değişkeni seçip, soldaki listeden sağdaki listeeye aktarın ve Options'u seçin. Açılan Options diyalog kutusunda ortalamalar, standart sapma, minimum ve maksimum değerler önceden seçilmişdir. Bunlara ek olarak seçebileceğiniz diğer istatistikler Frequencies diyalog kutusunun Statistics menüsünde de bulunan istatistiklerdir. Descriptives'de ortanca ve mod seçenekleri bulunmaz.



Şekil 7
SPSS'te
"Descriptives"
diyalog
kutusu.

Altgrupların tanımlayıcı istatistikleri

“Analyze >> Descriptive Statistics >” ya da “... >> Descriptives” menüleri ile tüm hastaların tanımlayıcı istatistikleri elde edilir. Oysa genellikle altgrupların tanımlayıcı istatistikleri merak edilmektedir. Örnek 29'daki çalışmada hasta ve normal deneklerin oluşturduğu heterojen grubun kan şekeri değerlerinin tanımlayıcı istatistikleri ile ilgilenilmediği açıklar. Asıl ilgilenilen konu, her altgrubun ayrı ayrı tanımlayıcı istatistikleridir.

Nominal ya da ordinal değişkenlerin her değeri için (örneğin grup adlı değişkenin değerleri tip I DM, tip II DM ve normal gruplardır) tanımlayıcı istatistikler farklı yöntemlerle elde edilebilir:

1. “Data >> Split file >> Compare Groups (ya da Organize Output by Groups)” seçeneği seçildikten sonra, *Groups based on...* listesine altgrupları belirleyen nominal ya da ordinal değişkenin adı seçilir. Eğer bu değişken değerleri sıralı değilse, kendiliğinden küçükten büyüğe sıraya konur ve ekranın sağ alt köşesinde *Split File On* yazısı çıkar. Bundan sonra yapılacak her türlü analiz, altgruplar için ayrı ayrı yapılır. Tüm grubun birlikte ele alınacağı grafik ya da analizler yapabilmek için Data >> Split files diyalog kutusuna yeniden gitip, *Analyze all cases, do not create groups* seçeneğini seçmek gerekir. Şekil 8'de kontrol, tip I DM ve tip II DM için sırasıyla 0, 1 ve 2 olarak kodlanmış olan tip adlı değişken, *Split files*'dan gruplayıcı değişken olarak seçilmiş ve daha sonra Frequencies >> Statistics'ten ortalama, standart sapma, minimum ve maksimum değerler seçilerek kreatininin tanımlayıcı istatistiği yapılmıştır.

Statistics				
KREATINI				
0 Kontrol	N	Valid Missing	15 0	
	Mean		1.271	
	Std. Deviation		.325	
	Minimum		.6	
	Maximum		1.7	
1 Tip I DM	N	Valid Missing	13 0	
	Mean		1.538	
	Std. Deviation		.739	
	Minimum		.5	
	Maximum		3.0	
2 Tip II DM	N	Valid Missing	11 0	
	Mean		1.186	
	Std. Deviation		.654	
	Minimum		.3	
	Maximum		2.5	

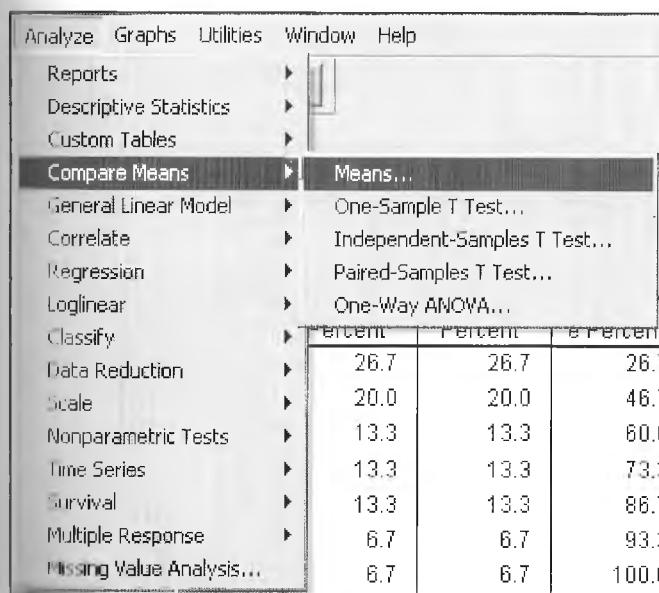
Şekil 8

Örnek 29'daki çalışmanın normal, tip I DM ve tip II DM gruplarının kreatininin değerlerine ilişkin bazı tanımlayıcı istatistikleri.

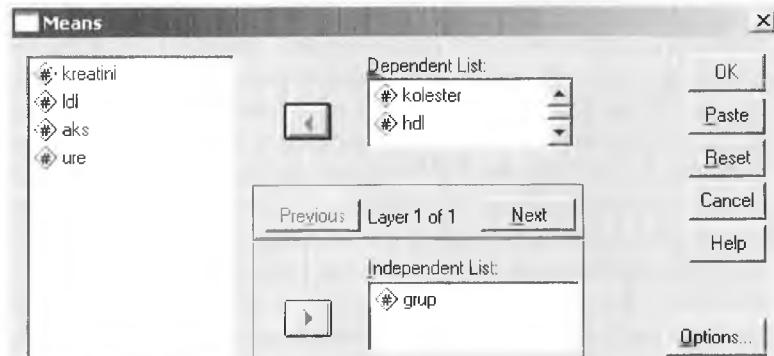
2. "Analyze >> Compare Means >> Means" menüsünde *Independent* listesine altgrupları belirleyen değişkenler, *Dependent* listesine ise tanımlayıcı istatistiği yapılacak değişkenler seçilir (Şekil 9 ve 10). *Independent* menüsüne birden çok değişken seçilebileceği gibi, tabakalandırma da yapılabilir. Örneğin *Independent* olarak hastalık grubu ve cinsiyet, *Dependent* olarak kreatinin alınırsa, tip I DM, tip II DM, kontrol grubu, erkekler ve kadınların kreatinin değerlerinin tanımlayıcı istatistikleri yapılır (Şekil 11). Her gruptaki farklı cinsiyetlerin kreatinin değerlerinin tanımlayıcı istatistiklerini elde etmek için *Independent* listesine önce hastalık grubunu seçip, *Next*'i seçin. Böylece hastalık grubu birinci tabakayı oluşturur. Daha sonra *Independent* listesine cinsiyet'i seçerek ikinci tabaka olarak, cinsiyet'i belirleyin. Böylece tip I DM erkekler, tip I DM kadınlar, tip II DM erkekler, tip II DM kadınlar, kontrol grubu erkekler ve kontrol grubu kadınlar olmak üzere altı altgrubun tanımlayıcı istatistiklerini elde edersiniz (Şekil 12).

"Analyze >> Compare Means >> Means" menüsündeki *Options*'da yapılabilecek çeşitli tanımlayıcı istatistikler vardır. Ayrıca birinci tabakadaki değişkenin belirlediği altgrup ortalamalarının karşılaştırıldığı tek yönlü varyans analizi (ANOVA) tablosu da yapılabilir.

3. "Analyze >> Descriptive Statistics >> Explore" menüsünden altgrupların tanımlayıcı istatistikleri elde edilebilir (Şekil 13 ve 14).



Şekil 9
SPSS'te "Compare Means >> Means" seçimi.



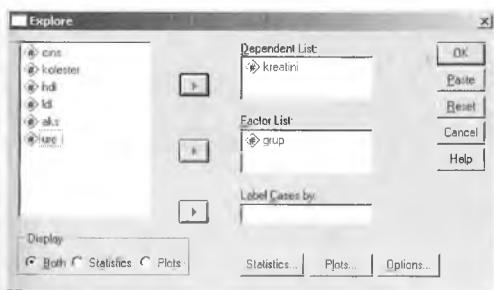
Şekil 10
SPSS'te
"Means"
diyalog
kutusu.

Şekil 11
Grup değişkeninin
belirlediği
altgrupların
kreatinin
değerlerinin
bazi tanımlayıcı
istatistikleri.

		Report			
		GRUP			
		0 Kontrol	1 Tip I DM	2 Tip II DM	Total
KREATINI	Mean	1.271	1.538	1.186	1 186
	N	15	13	11	11
	Std. Deviation	.325	.739	.654	.654

Şekil 12
Birinci tabaka olarak
hastalık grubu, ikinci
tabaka olarak
cinsiyetin seçildiği
altgrupların kreatinin
değerlerinin bazı
tanımlayıcı istatistikleri.

Report					
KREATINI					
GRUP	CINS	Mean	N	Std. Deviation	
0 Kontrol	1 Erkek	1.025	8	.191	
	2 Kadın	1.553	7	.174	
	Total	1.271	15	.325	
1 Tip I DM	1 Erkek	1.000	6	.276	
	2 Kadın	2.000	7	.700	
	Total	1.538	13	.739	
2 Tip II DM	1 Erkek	.717	6	.331	
	2 Kadın	1.749	5	.453	
	Total	1.186	11	.654	
Total	1 Erkek	.925	20	.286	
	2 Kadın	1.769	19	.508	
	Total	1.336	39	.588	



Sekil 13
SPSS'te "Explore" diyalog kutusu.

Descriptives				
	GRUP		Statistic	Std. Error
KREATINI	0 Kontrol	Mean	1.271	8.387E-02
		95% Confidence Interval for Mean	Lower Bound	1.092
			Upper Bound	1.451
		5% Trimmed Mean		1.286
		Median		1.200
		Variance		.106
		Std. Deviation		.325
		Minimum		.6
		Maximum		1.7
		Range		1.1
		Interquartile Range		.646
		Skewness		-.219
		Kurtosis		.580 -.511
	1 Tip I DM	Mean	1.538	.205
		95% Confidence Interval for Mean	Lower Bound	1.092
			Upper Bound	1.985
		5% Trimmed Mean		1.515
		Median		1.300
		Variance		.546
		Std. Deviation		.739
		Minimum		.5
		Maximum		3.0
		Range		2.5
		Interquartile Range		.650
		Skewness		1.123
		Kurtosis		.616 .971
	2 Tip II DM	Mean	1.186	.197
		95% Confidence Interval for Mean	Lower Bound	.747
			Upper Bound	1.625
		5% Trimmed Mean		1.162
		Median		1.200
		Variance		.428
		Std. Deviation		.654
		Minimum		.3
		Maximum		2.5
		Range		2.2
		Interquartile Range		.946
		Skewness		.542
		Kurtosis		.661 .098
				1.279

Sekil 14
Grup
değişkeninin
belirlediği
altgrupların
kreatinin
değerlerinin
tanımlayıcı
istatistikleri.

bölüm 6

ANALİTİK İSTATİSTİK

Güven aralığı ve güven sınırları

Hipotez testleri

H_0 ve H_1 , hipotezleri

P değeri ve yanılma düzeyi

Test istatistiği

Hipotezin yönü

Hipotez testinin hataları ve güç

Hipotez testi koşulları (ya da varsayımları)

Parametrik ve nonparametrik testler

Güç analizi ve örneklem büyüklüğünün saptanması

Tek grup ortalaması için örneklem büyüklüğü

Bilimsel çalışmaların amacı, örneklem değerlerinden evren değerlerinin kestirilmesidir. Evren parametrelerinin kestirilmesi için ya güven aralığı ve sınırları ya da hipotez testleri kullanılır.

GÜVEN ARALIĞI VE GÜVEN SINIRLARI

ÖRNEK 30

Multipl myelomada serum kalsiyum (Ca) düzeyinin sağlıklı kontrollerle karşılaştırıldığı bir çalışma正在被执行. Multipl myelomalı 81 hastada ortalama serum Ca düzeyi 11.5 mg/dL ve standart sapması 0.9 mg/dL ve 81 sağlıklı denek içeren kontrol grubunda ise ortalama serum Ca düzeyi 9.8 mg/dL ve standart sapması 0.9 mg/dL olarak bulunuyor. Çalışmanın amacı, iki grup arasında fark olup olmadığını ortaya konulması.

Myeloma grubundaki 81 hastanın Ca değerlerinin %95'i, ortalama \pm 2SD arasında yer alacaktır. Ortalama Ca değeri ise, %95 olasılıkla, ortalama \pm 2SEM arasında yer alacaktır. Yani "myelomalı hastalar" evreninden seçilecek 81 hasta içeren örneklemelerin %95'inde ortalama Ca değeri, bu grubun orta-

lama değeri $\pm 2\text{SEM}$ arasında yer alacaktır. Hesaplandığında bu aralığın alt sınırının $11.5 - 0.2 = 11.3 \text{ mg/dL}$ ve üst sınırının $11.5 + 0.2 = 11.7 \text{ mg/dL}$ olduğu görülmektedir. Kontrol grubunun Ca ortalaması olan 9.8 mg/dL ise $\%95$ olasılıkla $9.6 - 10.0 \text{ mg/dL}$ arasında yer alacaktır. İki grubun Ca ortalamalarının dahil oldukları bu değer aralıkları kesişmediği için, myeloma grubunun serum Ca düzeyi ortalamasının en azından $\%95$ olasılıkla normal kişilerdeki serum Ca düzeyinden farklı olduğu söylenebilir.

Bu örnekte myeloma grubunda serum Ca düzeyinin $\%95$ güven sınırları (*confidence limits*) 11.3 ve 11.7 mg/dL ve güven aralığı (*confidence interval*) $11.3 - 11.7 \text{ mg/dL}$ 'dir. $\%90$ ve $\%99$ güven sınırları da sıklıkla kullanılır. Yalnız ortalama değil, oran, relatif risk, "odds" oranı, bağıntı, iki ortalama arasında fark, iki oran arasındaki fark vb. bir çok istatistik için de güven sınırları belirlenebilir.

HİPOTEZ TESTLERİ

Örnek 30'da iki grup ortalamalarının birbirinden farklı olmadığı araştırılmış ve $\%95$ olasılıkla böyle olduğu anlaşılmıştır. Başka bir deyişle, myeloma grubu ve kontrol grubunun serum Ca düzeyi açısından aynı evrenden seçilip seçilmemiği test edilmiştir. Hipotez testleri de aynı amaca yönelik ve elde edilen sonuç açısından güven sınırlarıyla eşdeğer yöntemlerdir.

Istatistik hipotez testinin sorusu "İki grup, ortalama serum Ca değerleri açısından aynı evrenden mi seçilmiş?" dir ve bu sorunun yanıtının "evet" olduğu varsayılar. Hesaplamlar sonucunda bu soruya "evet" yanıtı verme olasılığı test edilir ve hesaplanan olasılık yeterince küçükse, hipotez kabul edilemez ve iki grubun aynı evrenden seçilmemiği, yani serum Ca düzeyi ortalamalarının farklı olduğu kabul edilir.

H_0 ve H_1 , hipotezleri

Farklılık olmadığını varsayıldığı hipoteze, yokluk hipotezi ya da farksızlık hipotezi (*null hypothesis*) adı verilir ve H_0 ile gösterilir. Yani H_0 hipotezi "... ile ... arasında fark yoktur" şeklindeki. H_1 ile gösterilen ve alternatif hipotez adlı verilen hipotez ise, H_1 hipotezinin tersidir. Yani H_1 hipotezi "... ile ... arasında fark vardır" şeklindeki. İstatistik analiz sonucunda H_0 hipotezi, yanı fark olmadığı reddedilince ne kadar yanlışlığımız hesaplanır. Örneğin hesaplanan olasılık $\%12$ ise, H_0 hipotezinin reddedilmesi ile $\%12$ olasılıkla yanlışlığıımız anlaşılr. H_0 hipotezi reddedilince H_1 hipotezinin doğru olduğu kabul edilmelidir. Eğer H_0 hipotezinin reddedilmesi ile yanlışılma olasılığı $\%12$ ise, H_1 hipotezinin kabul edilmesi ile yanlışlığının da $\%12$ olması gereklidir.

P değeri ve yanılma düzeyi

Örnek 30'daki sorunun çözümünde hipotez testi kullanılacaksa, H_0 hipotezi "myeloma grubu ile kontrol grubunun ortalama Ca değerleri aynıdır" ve H_1 hipotezi "myeloma grubu ile kontrol grubunun ortalama Ca değerleri farklıdır" şeklinde olmalıdır. Uygun hipotez testi ile hesaplanan olasılık 0.01, yani %1 olsun. Bunun anlamı nedir? Bunu aynı anlama gelen şu iki cümleyle açıklayalım: " H_0 hipotezini reddedersek ya da H_1 hipotezini kabul edersek, yanılma olasılığımız %1'dir"; yani "iki gruptaki Ca ortalamalarının farklı olduğunu %1 yanılma payı ile söyleyebiliriz". O halde "myelomada Ca ortalaması normalden farklıdır" yargısı %99 olasılıkla doğrudur.

Hesaplanan olasılık %1 değil de, %10 olsaydı, "myelomada Ca ortalaması normalden farklıdır" derken %90 olasılıkla doğru söylemiş olacaktık, ya da %10 olasılıkla yanlışmış olacaktık. Günlük konuşmalarda %90, %75, hatta %60 gibi olasılıkları çoğu zaman %100'e yakın anlamda kullanırız. Örneğin bir arkadaşınız "yarın saat 10'daki randevuya %90 gelirim" dediği zaman, bunu "yarın randevuya geleceğim"le aşağı yukarı eş anlamlı olarak algılar ve arkadaşınız randevuya gelmezse, şaşırırsınız. Aynı şekilde bir hastaya "bu ilaçla %80 iyileşirsiniz" dediğinizde eğer yanılırsanız, hastanız sizin doğru söylememekle suçlayabilir. Arkadaşınızın randevuya gelmemesi olasılığı olan %10 ve hastanın iyileşmemesi olasılığı olan %20'yi genellikle ihmali ederek, %0'mış gibi davranışır.

İstatistikte yanılma payı bu derece yüksek değildir. H_0 hipotezinin reddedilmesi için hesaplanan olasılığın %5 ya da daha az olması genellikle kabul edilen sınırdır; yani H_0 hipotezinin doğruluğu için hesaplanan olasılık %5 ya da daha küçükse, bu hipotezin kabul edilemeyeceği yargısına varılır. Bu değere yanılma düzeyi adı verilir ve α ile gösterilir. Hesaplanan olasılık ise, olasılık (probability) sözcüğünün ilk harfi olan p ile gösterilir. P değeri 0.05 olduğunda, H_0 hipotezinin tesadüfen doğru olma şansı %5'tir. "Myeloma ile kontrol gruplarının ortalama Ca değerlerinin farklı olmadığı" şeklindeki bir hipotezi reddederken eğer p değeri 0.07 ise, yalnız %7 olasılıkla yanılıyor, %93 olasılıkla doğru sonuca ulaşıyoruz demektir. Bu yanılma payı, günlük yaşamda kullandığımız anlamda düşünülürse, oldukça azdır. Ama istatistikte %7 olasılıkla yanilarak, bir hipotezi reddemezsiniz.

Hesaplanan p değerin 0.05'ten küçük olması durumunda H_0 hipotezi reddedilir. Örneğin p=0.001 ise H_0 hipotezi, yani iki grubun ortalama serum Ca düzeyleri arasında fark olmadığı varsayımlı %99.9 olasılıkla yanlıştır. P değeri 0.07 ise H_0 hipotezi %93 olasılıkla yanlıştır, ama doğru olduğu reddedilemez. Burada olduğu gibi p değeri 0.05'ten biraz büyük olduğunda H_0 hipotezinin olduğu gibi kabul edildiği anlamı çıkmamaktadır. Başka bir çalış-

mada farklılık saptanabilir. Ama p değeri 0.85 ise, H_0 hipotezinin yanlış olma olasılığı yalnızca %15'tir.

Bu nedenle p değerinin 0.05'ten büyük olması durumunda varılan sonuç H_0 hipotezinin doğru olduğu (yani fark olduğu) değil, doğru olduğunun reddedilemeyeceğidir.

Test istatistiği

Hipotez testlerinin hemen hemen hepsinde önce, p değerinin hesaplanmasıında kullanılacak bir rakam bulunur; daha sonra bu rakamın büyüklüğünde dayanılarak p değeri hesaplanır. Test istatistiği (*test statistics*) adı verilen ve t , Z , χ^2 , vb. adlarla anılan ve her durumda başka birinin kullanılmasının doğru olduğu rakamlardır. Neden kimi durumlarda t , kimi durumlarda χ^2 'nin kullanıldığı test edilen değişkenin dağılım özelliklerile ilişkilidir.

Hesaplanan test istatistiği, uygun tablolardaki değerlerle karşılaştırılır. Bu tablolarda sütunları yanılma düzeyi (α) ve satırları serbestlik derecesi (*degrees of freedom*) olarak gösteren rakamlar vardır. Genellikle her biri yanılma düzeyinin 0.50, 0.10, 0.05, 0.01, 0.001, ... gibi değerlerine karşılık gelen 5-6 sütun bulunur.

Serbestlik derecesi tanımı, her hipotez testi için farklıdır. Kimi testlerde denek sayıları, kimi testlerde grup sayıları, kimi testlerde de her ikisi birden serbestlik derecesinin hesaplanmasında kullanılır.

Hesaplanan test istatistiği, istenilen yanılma düzeyi ve o hipotez testinde belirtilmiş şekilde hesaplanan serbestlik derecesine karşılık gelen tablo değeri ile karşılaştırılır. Eğer hesaplanan test istatistiğinin mutlak değeri, tabloda ki değerden büyükse, α yanılma payı ile H_0 hipotezi reddedilir. Tablo 1'de "t tablosu"ndan bir örnek verilmiştir.

Örnek 30'daki çalışmada iki grup ortalamasının arasında fark olup olmadığını hipotez testi ile araştırırken, α yanılma düzeyini 0.05 olarak alalım. Bu durumda kullanılacak olan hipotez testinde t değeri hesaplanır.

Tablo 1 “0.10”, “0.05” ve “0.01” yanılma düzeyleri ve bazı serbestlik derecelerine karşılık gelen “t” değerlerinin gösterildiği, “t tablosu”ndan bir kesit.

Serbestlik derecesi	α yanılma düzeyi		
	0.10	0.05	0.01
1	6.314	12.706	63.657
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
10	1.812	2.228	3.169
15	1.753	2.131	2.947
20	1.725	2.086	2.845
30	1.697	2.042	2.750
40	1.684	2.021	2.704
60	1.671	2.000	2.660
120	1.658	1.980	2.617

Bu test için serbestlik derecesi iki gruptaki toplam denek sayısından 2 eksiktir. Toplam denek sayısı 162 olduğuna göre hesaplanan t değeri, tabloda yanılma düzeyi 0.05 olan sütunda, serbestlik derecesi 160 olan satırındaki değerle karşılaştırılacaktır. Tablodaki bu değer yaklaşık olarak 1.98'tir. Eğer hipotez testi ile hesaplanan t değerinin mutlak değeri 1.98'ten büyükse, iki grup ortalamasının farklı olmadığı hipotezi %5 yanılma payı ile reddedilebilir. Başka bir deyişle “myeloma grubunda ortalama Ca değeri, kontrol grubundan farklıdır; $p<0.05$ ” denebilir.

Makalelerde hesaplanan olasılıklar, yakın zamana kadar genellikle “ $p<0.05$, $p<0.001$, $p>0.05$ ” gibi değerlerle belirtilirdi. Bunun nedeni, bilgisayarların yaygınlaşmasından önce p değerlerinin tablolardan bulunması ve kesin bir değer yerine bazı değerlerden küçük ya da büyük olmasından dolayı net şekilde hesaplanamamasıdır. Artık bilgisayar yardımıyla kesin p değeri hesaplanabildiği için “ $p>0.05$ ” yerine “ $p=0.07$ ” ya da “ $p=0.85$ ” şeklinde belirtilmesi daha doğrudur. Çünkü p değerinin 0.07 ya da 0.85 olduğu her iki durumda da H_0 hipotezi reddedilememektedir, ama ikisinden çıkan anlamlı oldukça farklıdır.

Hipotezin yönü

H_0 hipotezi, yani alternatif hipotez, H_a hipotezinin tersidir. H_0 hipotezi “iki ortalama arasında fark yoktur” ise, H_1 hipotezi genellikle “iki ortalama arasında fark vardır” şeklinde belirtilir. Bu cümlede iki ortalamadan hangisinin diğerinden büyük olduğu belirtilmemiştir. Böyle yön belirtilmemiş bir alternatif hipoteze iki-yönlü (*two-tailed*) hipotez adı verilir. Çünkü fark olduğu takdirde, hangi ortalamanın daha büyük olduğu yönünde fikir belirtilmemiştir. Pratikte genellikle iki-yönlü hipotezler kurulur. Nadiren tek-yönlü hipotezler kullanılır. Örneğin yeni bir ilaçın, eski ilaçtan farklı olması yetmez, daha etkili olması gereklidir. Bu nedenle “yeni ilaçın iyileştirme oranı ile eski ilaçın iyileştirme oranları farklı mıdır?” sorusuna değil, “yeni ilaç daha iyi midir?” sorusuna yanıt ararız. Bu durumda kurulacak hipotez yön belirttiği için tek-yönlü hipotez olacaktır.

Hipotez testinin hataları ve güç

Teorik olarak hiç bir olayın gerçekleşme olasılığı %100'e, gerçekleşmemeye olasılığı ise %0'a ulaşamaz. Akşamki hava raporunda ertesi günün güneşli olacağını dinleyip de, en azından bir kez olsun yağmura yakalanmamış olan yoktur. Ağustos'ta kar yağmayaceği konusunda herkes iddiaya girebilir; kar yağmama olasılığı belki %99.999'dur, ama hiç bir zaman %100 değildir. Aynı şekilde istatistikte de hiç bir zaman %0 ve %100 olasılığa ulaşılabilir. Bir anlamda istatistik bilimi, bu gerçekten doğmuştur. Hipotez testleri ile ulaşılan sonuçlarda da %100 doğruluktan söz edilemez.

Hipotez testleri sırasında iki tür hata yapılabılır. Bu hatalardan biri yokluk hipotezinin doğru olduğu halde reddedilmesidir. Gerçekte myeloma grubunda ortalama Ca değeri ile kontrol grubunun ortalama Ca değeri arasında fark olmadığı halde, hipotez testi ile fark olduğu sonucu çıkabilir. α hatası ya da tip I hata olarak adlandırılan bu hata payı, genellikle 0.05 yani %5 olarak alınır. Bu durumda gerçekte fark olmadığı halde, %5 olasılıkla fark olduğu sonucuna varılması kabul edilir.

Tip I hatanın tersine gerçekte fark olduğu halde test sonucunda fark olmadığı sonucuna da varılabilir. Myeloma grubunun ortalama Ca değeri, kontrol grubundan daha yüksek olduğu halde hipotez testi sonucunda iki grubun ortalamalarının farklı olmadığı sonucu çıkabilir. Bu hataya da β hatası ya da tip II hata adı verilir. Tip I hata yalancı pozitiflik, tip II hata ise yalancı negatifliğe eşdeğer olarak düşünülebilir.

Gerçekte fark olduğu zaman test sonucunda da fark olduğu sonucuna varılması olasılığına ise testin gücü (*power*) adı verilir. Güç, $1-\beta$ 'ya eşittir. Myelomada Ca değerinin normalden daha yüksek olduğu gerektiğini %90 olasılık-

la doğru olarak saptayan bir testin gücü 0.90'dır.

Bir testte tip I hata ne kadar küçük tutulursa, tip II hata o kadar artar. Örneğin, test sonucunda p değerinin 0.05'ten küçük olması durumunda "fark vardır" yorumunu kabul etmek, %5 ya da daha az olasılıkla aslında fark olmadığı halde bu sonuca varmak gibi bir hatayı göze almak demektir (tip I hata). Eğer bu hatayı daha az olasılıkla, örneğin %1 olasılıkla göze almak istiyorsanız, bunu ancak gerçekte fark olduğu halde bunu gözden kaçırma olasılığının (tip II hata) artması pahasına yapabilirsiniz. Tip II hata olasılığı artınca, testin gücünü azaltacaktır. Testin gücünü azaltmadan her iki tip hatayı da azaltmak, ya da bir tip hatayı azaltırken, diğer tip hatanın artmamasını sağlamak için denek sayısı artırılmalıdır.

Hipotez testi koşulları (ya da varsayımları)

Istatistik analizi (1) uygun yöntemin seçilmesi, (2) yönteme özgü hesaplamaların yapılması ve (3) sonuçların yorumlanması olarak üç aşamada ele alınarak, en kolay aşama hesaplamalarıdır; çünkü artık güçlü ve hızlı bilgisayarlar ve kolay kullanımlı istatistik yazılımlar sayesinde en karmaşık hesaplamalar bile saniyeler içerisinde yapılabilmektedir. Her türlü veriye, tüm hipotez testlerindeki ya da güven sınırlarının hesaplanmasında kullanılan formüller uygulanarak, matematik işlemler yapıldığında bir test istatistiği ve buna karşılık gelen bir p değeri ya da güven sınırları bulunabilir. Ama istatistik analizde başta gelen sorun, örneklemden elde edilen sonuçların evrene ekstrapole edilebilirliğini sağlamaktır. Bu nedenle istatistik analizde asıl sorun doğru yöntemin seçilmesi ile ilgilidir. Uygun olmayan yöntemlerle yapılan analiz sonuçlarının evreni temsil ettiğini öne sürmek, bilimsel olarak hatalı olmasının yanında, insan yaşamı ile ilgili taşıdığı sorumluluk nedeniyle tip alanında daha da önemlidir. Çünkü bir ilaçın etkinliği, bir tanı yönteminin yararı gibi konularda tıbbi dergi editörlerinin de gözünden kaçabilecek hatalarla yayınlanmış bir makaledeki sonuçlara göre klinik yaklaşımını değiştirecek bir araştırmacı, dolaylı olarak hastalarını riske atmış olacaktır. Bu nedenle sağlık alanında çalışanların hem araştırma yaparken, hem de makale okurken, istatistik analizle ilgili temelere özel önem vermesi kaçınılmazdır.

Her hipotez testi, her türlü veriye uygulanamaz. Hipotez testlerinin uygulanabilmesi için verilerin dağılım ve yayılma özellikleri, denek sayısı, çalışmadaki grup sayısı, değişkenlerin ve grupların bağımlı ya da bağımsız olması gibi bazı koşulların sağlanması, diğer bir tanımla hipotez testinin uygulanabilmesi için gerekli varsayımların yerine getirilmesi gereklidir. Hangi hipotez testi için ne gibi koşulların arandığı ilgili bölümlerde anlatılmıştır. Bu koşulların bazlarının sağlandığını anlamak kolaydır. Örneğin tek-yönlü var-

yans analizinin uygulanabilmesi için grupların bağımsız olması, yani bir grupta yer alan bir deneğin aynı zamanda başka bir grupta da bulunmaması gereklidir. Örneğin diyabetik hastalarda normoglisemik ve hiperglisemik dönemde alınan kan örneklerinde fruktozamin düzeyi ölçülmüş ve sağlıklı kontrollerle karşılaştırılmak isteniyorsa, bu çalışmada bulunan bu üç gruptan birinde yer alan bir deneğin aynı zamanda başka bir grupta da yer almaması gereklidir. Bir hastanın hem hiperglisemik, hem de normoglisemik dönemdeki fruktozamin değerinin analize alınan gruplarda bulunması hatalı olur. Çalışmadaki verilerin bu koşula uygun olup olmadığını anlamak için bir test uygulamaya gerek olmadığı açıktır.

Öte yandan hipotez testinin uygulanabilmesi için gereken bazı koşulların sağlanıp sağlanmadığını anlamak her zaman bu kadar kolay olmayabilir. Örneğin bir çok hipotez testinin koşullarının başında gelen verilerin normal dağılıma uyup uymadığını ancak uygun grafik ya da analiz yöntemleri ile saptayabiliyoruz.

Istatistik analize başlamadan önce çalışma düzenine ve verilere uygun hipotez testlerinden hangisi ya da hangilerinin uygulanabileceğine karar vermek için mutlaka ilgili testlerin koşullarının sağlanıp sağlanmadığı araştırılmalıdır. Buna dikkat edilmeme sebebi ne olur? Birincisi, yokluk hipotezi yanlış olmadığı halde reddedilebilir. Örneğin iki gruba ait verilerin ortalamaları arasında fark olup olmadığı araştırılıyorsa, yokluk hipotezi “İki grup ortalaması eşittir” şeklinde olacaktır. Eğer uygun koşullar sağlanmadığı halde hipotez testi uygulanırsa, eldeki veriler, iki grup ortalaması arasında fark olmadığını gösterse bile fark varmış gibi sonuç elde edilebilir. Bazen de tersine, gerçekte fark olduğu halde, uygun koşullar sağlanmazsa, fark olmadığı sonucuna varılabilir.

Parametrik ve nonparametrik testler

Çalışma düzenine ve verilere uygun istatistik analiz yöntemlerinden hangisinin seçileceğine karar vermek için test koşullarının sağlanması gereğinden söz ettik. Aynı durum için uygun olan seçeneklerden bazlarında istenen koşulları sağlamak pek kolay olmamaktır ya da verilerin dağılım özgürlüğü nedeniyle uygulanması sakıncalı olabilir. Bu durumda daha az koşulun sağlanması durumunda da uygulanabilen seçenekleri uygulamak daha yerinde olacaktır. Nonparametrik testler, daha az koşulun varlığında uygulanabilen testlerdir. Parametrik testlerin hemen hemen hepsinin uygulanabilmesi için ise, en azından verilerin normal dağılıma uyması ve her testte farklı olmak üzere başka koşulların da sağlanması gereklidir. Parametrik ya da nonparametrik testlerin genel bir tanımı yoktur. Genel olarak nonparametrik testlerle nominal, ordinal ya da normal dışı dağılım gösteren sayısal veriler analiz edilebilirken, parametrik testlerle normal dağılım gösteren sayısal verile-

rin analizi yapılabilir. Öte yandan normal dağılıma uygun verilere nonparametrik test uygulanması pek hatalı sayılmazken, ordinal ya da normal dışı dağılım gösteren verilere parametrik test uygulanması daha sakıncalıdır.

Her testin uygulanabilmesi için gerekli koşulların neler olduğu ve verilerin bu koşullara uygunluğunun nasıl saptanacağının mutlaka iyi bilinmesi gereklidir. Eğer koşulların sağlanıp sağlanmadığını bilmiyorsanız, verilerin analizinde nonparametrik test kullanılması daha güvenli olur. Ancak parametrik test için gerekli koşulların sağlanmasına karşın nonparametrik test uygulanması halinde, parametrik testlerin kendine özgü avantajlarından yararlanmamış olursunuz.

GÜC ANALİZİ VE ÖRNEKLEM BÜYÜKLÜĞÜNÜN SAPTANMASI

Hipotez testinin hataları ve gücü, bir çalışmaya alınacak denek sayısını, yani örneklem büyüklüğünü belirlemek için kullanılabilir.

ÖRNEK 31

Bir antihipertansif ilacın sistolik kan basıncını 20 mmHg kadar düşürüp düşürmediğini araştırmak için yapılan bir çalışma 20 hasta üzerinde yapılrsa, çalışmada kullanılan hipotez testlerinin gücü, 50 hasta kullanıldığı bir çalışmaya göre daha az olacaktır. Bu nedenle gerçekte ilacın etkisi olduğu halde, etkisi yokmuş gibi hatalı (tip II hata) bir sonuca ulaşılabilir. Öte yandan 50 hasta yerine 250 hasta kullanılması ile çalışmanın gücü daha da artacaktır. Bu durumda da 50 ile 250 hastanın kullanılması durumunda çalışmanın masrafları artacak, ve belki de tip II hatadaki azalma çok az (örneğin %8'den %6'ya azalma gibi) olacaktır.

Hipotez testi ile hesaplanan test istatistiğinin tablodaki hangi değerle karşılaştırılacağını seçerken serbestlik derecesinin hesaplanması denek sayısının belirleyici olduğunu söz edilmiştir.

Örneğin 81'er denekli bağımsız iki grubun ortalamalarının karşılaşılmasıında hesaplanan t değerinin 2.02 olduğunu varsayıyalım. 160 (81+81-2) serbestlik derecesinde ve 0.05 yanılma düzeyindeki tablo t değeri 1.98 olduğu için, iki grubun ortalamalarının birbirinden farklı olduğu %5 yanılma payı ile söylenebilir. Oysa gruplar 81'er değil de 15'er denekten oluşsaydı, 28 (15+15-2) serbestlik derecesindeki t değeri 2.05 olduğu için ortalamaların farklı olduğu söylenenemeyecekti.

Denek sayısı arttıkça tablodan seçilecek t değerine karşılık gelen serbestlik derecesi de artar. Serbestlik derecesi arttıkça tablodaki t değeri küçülür. Bu nedenle denek sayısı arttıkça, t testi ile hesaplanan daha küçük t değerle-

ri için bile, karşılaşılacak tablo t değeri de küçüleceği için, ortalamaların farklı bulunabilmesi şansı, yani testin gücü artar.

Denek sayısı azaldıkça, testin gücü azalır, yani β hatası artar. Bu nedenle, çalışmaya başlamadan önce optimum denek sayısının belirlenmesi giderek daha çok üzerinde durulan bir konu olmaya başlamıştır.

Her istatistik yöntemin gücü ve β hatası kendine özgüdür. Bu nedenle örneklem büyülüğü farklı çalışma düzenleri için farklı formüllerle hesaplanır. Bu kitapta örnek olarak tek grup ortalamasının değerlendirileceği bir çalışmada örneklem büyülüğünün hesaplanması anlatılacaktır. Güç analizi ve örneklem büyülüğünün saptanması sırasında mutlaka bir istatistik uzmanına danışılması gereklidir.

Tek grup ortalaması için örneklem büyülüğü

Örneklem büyülüğünü belirlemek için, çalışmaya başlamadan önce bazı kararlar vermek gereklidir.

Bunlar;

1. Çalışma grubunun ortalaması, karşılaşılacak değerden farklı olduğu halde, testte bu farkın saptanamaması olasılığı (yani α yanılma düzeyi) ne kadar olmalıdır? α yanılma düzeyi genellikle 0.05 olarak seçilir.
2. Çalışma grubunun ortalaması, karşılaşılacak değerden farklı ise, bunun saptanabilmesi olasılığı (yani testin gücü) ne kadar olmalıdır? Testin gücü genellikle 0.95 ile 0.80 arasında seçilir. Buna karşılık gelen β hatası değeri 0.05 ile 0.20'dir.
3. Çalışma grubunun ortalaması ile karşılaşılacak değer arasındaki ne kadar fark önemlidir ya da saptanmalıdır? Bu karar, her durumda farklı olabilir. Örneğin myelomada serum kalsiyum (Ca) düzeyinin normal değerlerin üst sınırından yüksek olup olmadığı araştırma konusu ise, farkın 2 mg / dL olması önemli olan en az fark olarak seçilebilir. Eğer, araştırmanın konusu myelomada serum Ca değerlerinin, normal değerler içine girse bile, yüksek olup olmadığı ise, 0.2 mg / dL fark bile önemli olabilir. Bunun seçimi tamamen araştırıcıya kalmıştır.
4. Karşılaştırılacak standart değerin (yani evren ortalamasının) standart sapması ne kadardır? Bu soruya her zaman yanıt bulunamayabilir. Çünkü çoğu zaman evren ortalamasının standart sapması bilinemez. Bu durumda gerçeğe yakın olabilecek bir kestirim yapılabilir. Genellikle bilinen, standart bir değerin alt ve üst sınırları ortalama $\pm 3SD$ ile saptanır. Örneğin kadınlardaki normal Hb değerleri 12-16 g / dL ise, ortalama Hb değerinin 14 g / dL ve standart sapmanın ise $2 / 3 = 0.7$ g / dL olduğu varsayılabılır. Bu yalnızca bir kestirimdir. Çok uzun süredir ya da yaygın ola-

rák kullanılmayan değerlerin alt ve üst sınırları ile standart sapmayı kesitmek yaniltıcı olabilir. Ama ne olursa olsun, örneklem büyülüğünü hesaplayabilmek için evren ortalamasının standart sapması bilinmeli, ya da en azından kestirilmelidir.

Bu dört soruya yanıt bulunduktan sonra tek grup ortalaması için örneklem büyülüğü aşağıdaki formülle hesaplanır.

$$n = \left[\frac{(Z_\alpha - Z_\beta)\sigma}{\mu_1 - \mu_0} \right]^2$$

Bu formülde σ (sigma) evrenin standart sapması ve μ_1 ve μ_0 çalışma grubu ve evren ortalamalarıdır. $\alpha = 0.05$ için $Z_\alpha = 1.96$ ve $\beta = 0.10$ için $Z_\beta = -1.28$ 'dır.

ÖRNEK 32

Myelomada serum Ca düzeyi ile normal Ca değeri arasında en az 0.2 mg/dL fark olmasının önemli olduğu ve normal kalsiyum değerinin standart sapmasını 0.3 mg/dL olduğu varsayımları ve $\alpha = 0.05$ ve $\beta = 0.10$ yanılma payları ile çalışmaya alınması gereken hasta sayısı en az kaç olmalıdır?

Örnek 32'deki değerleri yukarıdaki formülde yerine koyarsak;

$$n = [(1.96 + 1.28)(0.3) / 0.2]^2 = 23.6$$

O halde, çalışmaya en az 24 hasta alınmalıdır. Formülle hesaplanan değer tam sayı değilse, her zaman bir üstteki sayı hasta sayısı olarak alınmalıdır. Örneğin hesaplanan değer 23.05 bile olsa, örneklem büyülüğü olarak 24 seçilmelidir.

Eğer aradaki farkın 0.4 mg / dL olması durumu en az önemli fark olarak seçilirse, yani fark 0.4 mg / dL'den küçük olduğunda bunun önemli olmadığı gibi bir klinik yargı varsa, hesaplanan hasta sayısı 6 olacaktır.

Çalışmaya başlamadan önce yapılacak güç analizi ile gerçek farkların gözden kaçabilecegi yetersiz sayıda hasta kullanılması ya da çalışma maliyetini gereksiz yere artıracak haddinden fazla hasta kullanılması gibi hatalar önlenebilir.

bölüm 7

SAYISAL VERİLER İÇİN TANIMLAYICI GRAFİK VE TABLOLAR

Histogram

Gövde-yaprak grafiği

Boksör torbası (kutu-nokta) grafiği

Sıklık tabloları

SPSS'te sıklık tablosu ve tanımlayıcı grafikler nasıl yapılır?

Frequencies menüsü

Charts altmenüsü

Explore menüsü

Bir çalışmada elde edilen verilerin tablolar ve grafikler şeklinde sunulması, en etkili anlatım yöntemidir. Bazan birkaç paragrafta anlatılan veriler, uygun bir tablo ya da grafikte özelendiği zaman, daha kolay anlaşılır hale gelir. Çalışmada elde edilen verilerin türüne göre kullanılması uygun olan grafik türü farklıdır.

Sıklık tabloları ve çapraz tablolar, gruplandırılabilen değişkenlere ait verilerin özetlenmesi için yararlıdır, ama sayısal veriler bu tablolar ile gösterilemez. Örneğin tip I diabetes mellitus (DM), tip II DM ve normal kontrol grublarındaki hasta sayıları ve yüzdeleri sıklık tabloları ve çapraz tablolar ile belirtilirken, her gruptaki trigliserid değerlerinin ortalaması, standart sapma, ortanca, en küçük ve en büyük değeri, ... gibi değişkene ait veriler sıklık tablosu ile gösterilemez. Hangi tanımlayıcı istatistik değerlerinin verilere uygun olduğunu anlamak için verilerin ölçüm düzeylerinin bilinmesi gereklidir.

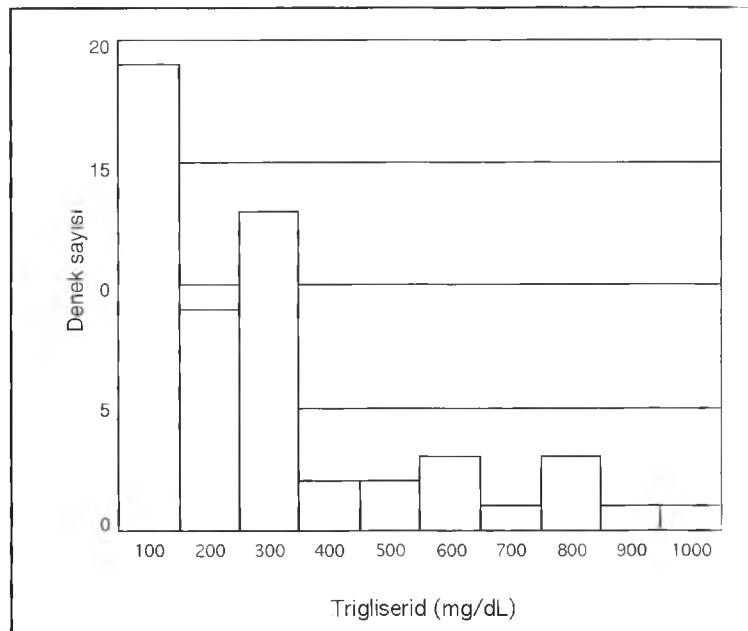
Istatistik analize başlamadan önce, değişkenlerin aldığı değerlerin kabaca gözden geçirilmesi, hem hatalı veri girişi varsa, onları saptamak, hem de uygun istatistik analiz yöntemlerinin seçilmesi için gereklidir. Bu amaçla değişkenlere ait bazı grafik ve tabloların incelenmesi yararlıdır. Bu grafiklerin başında gelen histogram, gövde-yaprak grafiği, boksör torbası grafiği ve sıklık tabloları ile ilgili bazı ayrıntıları öğrenmek ve bu grafik ve tabloları değerlendirebilmek gereklidir.

HİSTOGRAM

Yatay eksende değişken değerlerinin, dikey eksende ise denek sayısının yer aldığı ve her değişken değerinin gözlendiği denek sayısının çubuklar şeklinde gösterildiği grafiğe histogram denir.

Histogramda değişken değerleri eşit aralıklara bölünüp yatay eksende, her aralığa düşen denek sayısı ise dikey eksende belirtilir. Histogramdaki her çubuk, karşılık geldiği aralığa düşen denek sayısını gösterir. Yatay eksende her çubuğa karşılık olarak, çubugun temsil ettiği aralığın orta değeri yazılır.

Şekil 1'de SPSS'ten çizdirilmiş tip II DM hastalarında trigliserid değerlerinin histogramı görülmektedir. Bu şekilde birinci çubugun temsil ettiği aralığın orta değerinin 100, ikinci çubugun 200, ... ve son çubugun 1000 olduğu görülmektedir. Trigliserid değerleri 50-150 arasında olan 19 hasta olduğu anlaşılmaktadır. Dikey eksen, ölçülen değişkenin her aralığına düşen hasta sayısı yerine o aralıktaki hastaların toplam hasta sayısına oranı olarak da tanımlanabilir. Her iki halde de histogramın şekli değişmez.



Şekil 1
Tip II DM hastalarında trigliserid değerlerinin dağılımını gösteren histogram.

GÖVDE-YAPRAK GRAFIĞI

Gövde-yaprak grafiği (*stem-and-leaf plot*), histograma göre görsel olarak daha az anlaşılsa da, değerler hakkında daha fazla bilgi verir.

Gövde-yaprak grafiği, çubukların yatay olarak yer aldığı histogram gibidir. Her satırın uzunluğu, o satıra karşılık gelen değer aralığındaki denek sayısını temsil eder. Ama gövde-yaprak grafiğinde her deneğin gerçek değeri görülür. Gövde-yaprak grafiği çizmek için, her deneğin değeri en sağdaki rakam (yaprak) ve diğer rakamlar (gövde) olarak ikiye ayrılır. Örneğin 173 rakamında "17" gövdede yer alırken, "3" rakamı gövdeye yapışık bir yaprağı temsil edecektir.

Şekil 2 ve 3'de SPSS'ten elde edilmiş gövde-yaprak grafikleri görülmektedir.

Şekil 2'de *Stem* sütununda değerlerin gövdesi, *Leaf* sütununda yaprakları gösterilmektedir. *Frequency* sütununda her gövde parçasına (değer aralığına) düşen yaprak (denek) sayısı gösterilmektedir. Denek değerleri küçükten büyüğe sıralandığında 17, 17, 17, 18, 18, ..., 59, 68 olarak görülmektedir.

Yaş Stem-and-Leaf Plot			
Frequency	Stem	&	Leaf
9.00	1	.	777888999
14.00	2	.	11234456677889
11.00	3	.	23355666678
12.00	4	.	122222333568
7.00	5	.	0234679
1.00	6	.	8
Stem width:		10	
Each leaf:		1 case(s)	

Şekil 2
Yaş
değişkeninin
gövde-yaprak
grafiği.

Gövde genişliği (*Stem width*) gövdedeki rakamların hangi basamağa ait olduğunu göstermektedir. Bu örnekte gövde genişliği ondalık basamak olduğu için *Stem width* 10 olarak belirtilmiştir. Hasta değerleri 99'dan büyük değerler de içeriyorsa (triglycerid gibi), gövde genişliği 100 olacaktır. Bu durumda her hastayı temsil eden yaprak ise, hastanın değerinin noktanın solundaki ikinci rakam olarak yazılır. Bu durumda değerler yuvarlanmış olur. Örneğin, Şekil 3'de triglycerid değeri 632 mg/dL olan bir hasta "6."

nolu gövde parçasında "3" yaprağıyla gösterilmiştir, yani 630'a yuvarlanmıştır. Şekil 3'de trigliserid değeri aşırı ucta ($\geq 670 \text{ mg/dL}$) yer alan altı hasta *Extremes* satırında gösterilmiştir.

Şekil 3
Trigliserid
değişkeninin
gövde-
yaprak
grafiği.

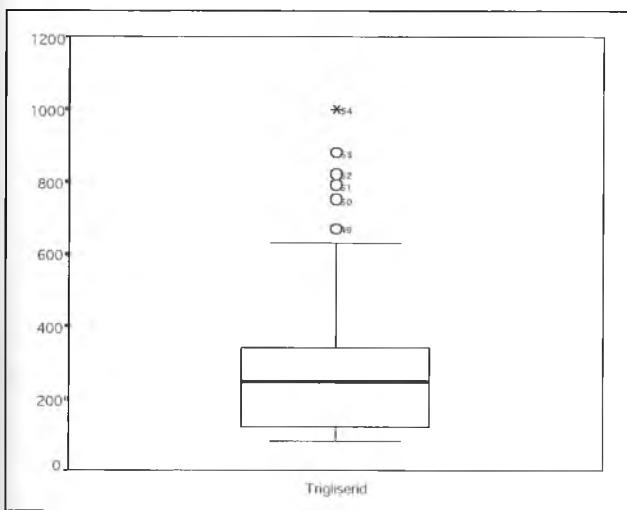
Trigliserid Stem-and-Leaf Plot					
Frequency	Stem	&	Leaf		
6.00	0	.	788889		
16.00	1	.	0111112222334568		
12.00	2	.	123444566888		
8.00	3	.	22233345		
1.00	4	.	0		
4.00	5	.	0467		
1.00	6	.	3		
6.00	Extremes		(≥ 670)		
Stem width: 100					
Each leaf: 1 case(s)					

BOKSÖR TORBASI (KUTU-NOKTA) GRAFİĞİ

Bir diğer grafik türü olan boksör torbası ya da kutu-nokta grafiği (*box-plot*), verilerin tanımlayıcı istatistiğini özetleyen oldukça yararlı bir gösterim şeklidir. Bu tür grafikte tek tek değerler yerine dağılımla ilgili bazı istatistikler yer alır. Bunlar ortanca ve çeyrekler ve bu değerlerden çok uzak olan aşırı değerlerdir.

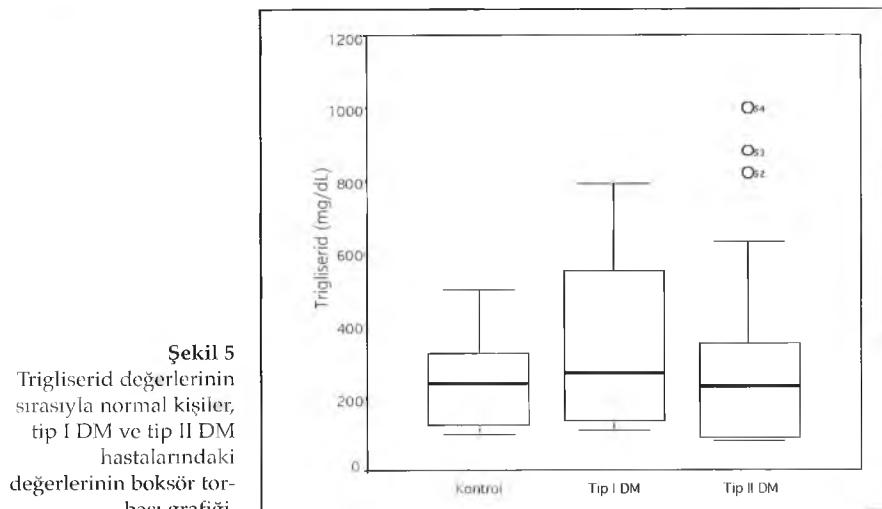
Şekil 4'te trigliserid değerlerinin SPSS'ten çizdirilmiş boksör torbası grafiği görülmektedir. Kutunun alt sınırı 25. persentil, üst sınırı 75. persentili, kutunun içindeki yatay çizgi ortancayı temsil etmektedir. Kutunun yüksekliği çeyreklerarası aralığa (*interquartile range*) karşılık gelmektedir. Boksör torbası grafiğinde aşırı değerler de gösterilir. SPSS'te kutuların alt ya da üst sınırlarından 3 kutu boyu ya da daha uzaklıkta olan değerler asteriks (*), 1.5-3 kutu boyu uzaklıkta olan değerler içi boş daire (O) ile gösterilmektedir. Bu işaretlerin yanındaki rakamlar hangi deneye ait olduğunu göstermektedir. Asteriks ile işaretli değerlere uç değerler (*extreme values*), daire ile işaretli değerlere taşan değerler (*outliers*) adı verilir. Kutuların alt ve üst kenarlarının

ortasından çizilen dikey çizgiler ise, *outlier* olmayan, yani alt ya da üst kenarlar uzaklığı 1.5 kutu boyundan az olan en büyük ve en küçük değerlere kadar uzanmaktadır.



Şekil 4
Trigliserid
değerlerinin boksör
torbası grafiği.

Boksör torbası grafiğine bakılarak ne söylenebilir? Ortanca yerine bakarak merkezi eğilim hakkında fikir edinilebilir. Kutunun uzunluğundan değerlerin dağılımı hakkında fikir edinilebilir. Ortanca, kutunun ortasında değilse, dağılımin çarpık olduğu söylenebilir. Eğer ortanca, kutunun alt kenarına yakınsa, sağa çarpık, üst kenarına yakınsa sola çarpık dağılım olduğu söylenebilir. Birden çok grubun değerlerinin dağılımını karşılaştırmak için boksör torbaşı grafiği görsel olarak çok yararlıdır (Şekil 5).



Şekil 5
Triglycerid değerlerinin sırasıyla normal kişiler, tip I DM ve tip II DM hastalarındaki değerlerinin boksör torbası grafiği.

SIKLIK TABLOLARI

Bazen sayısal değişkenlerin ortalama, ortanca, standart sapma vb. tanımlayıcı istatistikleri yanında sıklik tabloları (*frequency tables*) adı verilen tablolarda özetlenmesi de yararlı olabilir.

ÖRNEK 33

Bir çalışmada 200 kişinin ortalama Hb değeri 14.0 g/dL, standart sapması 1.6 g/dL, minimum ve maksimum değerleri 11.0 g/dL ve 17.8 g/dL olarak bulunuyor. Hb değerlerinin tanımlayıcı istatistikleri yanında 1 g/dL'lik gruplara bölüp, 11.0-11.9, 12.0-12.9, ..., 17.0-17.9 gruplarına giren kişilerin sayı ve yüzdeleri de verilmek isteniyor.

Örnek 33'te yapılmak istenen tablo, Hb değerlerinin sıklik tablosudur. Sıklık tablosu aşağıdaki adımlara göre hazırlanmalıdır:

1. Minimum ve maksimum değerleri belirleyin (Minimum: 11.0 ve maksimum: 17.8).
2. Maksimumdan minimumu çıkarıp, değer aralığını bulun ($17.8 - 11.0 = 6.8$).
3. Denekleri kaç sınıfa ayıracagınızı belirleyin. Genellikle 5-15 arasında

sınıfa ayırmak uygundur. Daha az sınıfı ayırmak ile ortalama, standart sapma ve değer aralığının verdikleri dışında çok fazla bilgi verici bir tablo olmaz. 15'den fazla sınıf olduğu zamanda tablo karışık olur (Yukarıdaki örnek için 5-7 sınıf yeterli olabilir).

4. Sınıf aralığını belirlemek için değer aralığını, sınıf sayısına bölün. Sınıf aralığı, sınıfı belirleyen en küçük ve en büyük değerler arasındaki farktır ($6.8/7 = 1$).
- a. Sınıf sınırları kesişmemelidir. Örneğin 30-39, 40-49 gibi sınıf sınırları tanımlanabilir. 30-40, 40-50 gibi tanımlanmış sınıflarda, bir alt sınıfın üst sınırı ile bir üst sınıfın alt sınırı karşılaşacağı için yanlış olur.
- b. Sınıf genişliği eşit olmalıdır. Örneğin bir sınıfın 30-39, diğer sınıfın 40-59 olması doğru olmaz.
- c. Sınıflar açık uçlu olarak (49 ve daha büyük gibi) tanımlanmamalıdır. Örnek 33'teki Hb değerleri Tablo 1'deki sıkılık tablosunda özetlenmiştir.

Tablo 1 Örnek 33'teki Hb değerlerinin sıkılık tablosu.

Hb (g/dL)	Sayı	Yüzde	Kümülatif yüzde
11.0-11.9	25	12.5	12.5
12.0-12.9	25	12.5	25.0
13.0-13.9	32	16.0	41.0
14.0-14.9	68	34.0	75.0
15.0-15.9	25	12.5	87.5
16.0-16.9	15	7.5	95.0
17.0-17.9	10	5.0	100.0
Toplam	200	100.0	

SPSS'TE SIKLIK TABLOSU VE TANIMLAYICI GRAFİKLER NASIL YAPILIR?

Frequencies menüsü

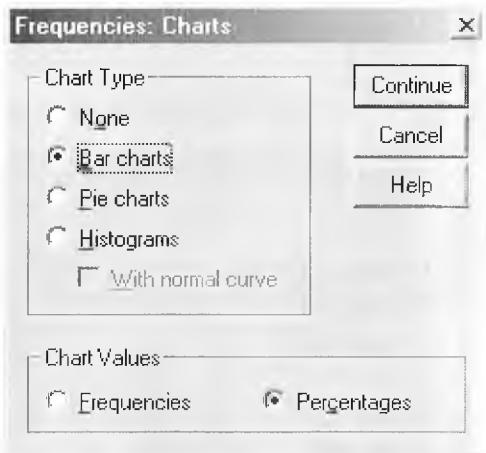
ÖRNEK 34

Bir çalışmada kronik aktif hepatit (KAH), siroz, hepatosellüler kanser (HCC), bilier siroz ve karaciğer dışı hastalıklara bağlı asiti olan kişiler olmak üzere toplam 48 hastada asit sıvısında kalsiyum (Ca), demir (Fe), bakır (Cu), çinko (Zn) ve magnezyum (Mg) düzeylerine bakılıyor.

Bu çalışmaya ait sıklık tablolarını ve histogramı elde etmek için **Frequencies** seçeneği seçilmelidir. **Frequencies** menüsü ile nasıl çalışılacağı Bölüm 5'te anlatılmıştır.

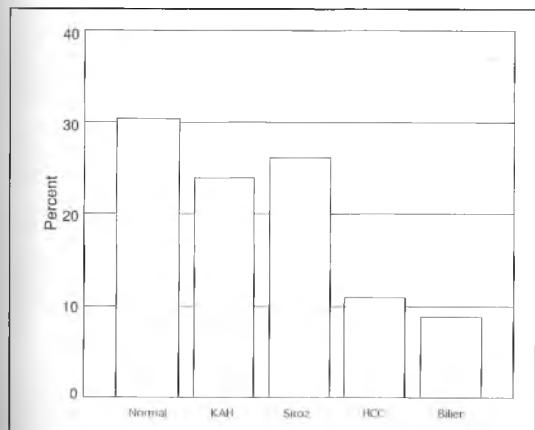
Charts altmenüsü

Bar grafik ya da histogram elde etmek için, *Charts*'ı seçin (Şekil 6). Grafik çizilmemesini belirten *None* önceden seçilmiştir. *Bar chart* seçilirse, yatay eksende değişkenin kategorileri, dikey eksende ise kategoriye düşen denek sayısı ya da yüzdesi bulunan çubuk grafik çizilir. *Chart Values* (Grafik Değerleri)'nde *Frequencies* seçilirse, dikey eksende denek sayısı, *Percentages* seçilirse, yüzde bulunur. Çubuk grafikte çubuklar bitişik değildir.



Şekil 6
SPSS'te "Frequencies:
Charts" diyalog kutusu.

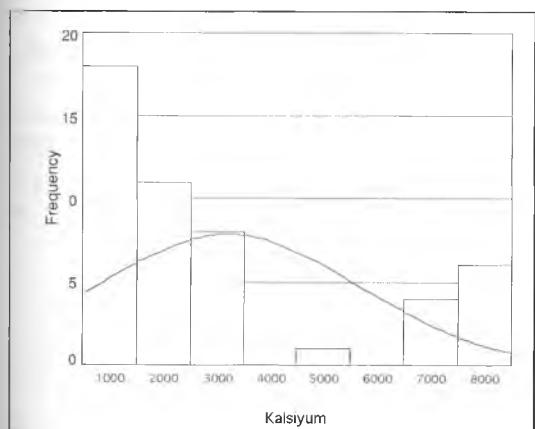
Şekil 7'de hastaların gruplara göre dağılımı yüzde olarak çubuk grafikle gösterilmiştir. Bu grafik için *Chart type* menüsünden *Bar chart* ve *Chart Values* menüsünden *Percentages* seçilmiştir.



Şekil 7
Örnek 34'teki çalışmada hastaların grplara göre dağılımı (Yüzde olarak).

Histogram seçeneği yalnız sayısal değişkenler için mümkündür. İstenirse, histogramın üzerine teorik normal dağılım eğrisi süperpoze olarak çizdirilebilir. Bunun için *with normal curve* seçilmelidir. Histogramdaki çubuklar bitişik olarak çizilir. Histogramdaki çubukların tepe noktalarının dağılımı, normal dağılım eğrisi ile aşağı yukarı çakışısa, değişken değerlerinin normal dağılım gösterdiği varsayılabılır.

Şekil 8'de normal grupta Ca değerlerinin histogramı ve beklenen teorik normal dağılım eğrisi görülmektedir. Bu grafik için *Chart type* menüsünden *Histogram* ve *with normal curve* seçilmiştir. Dağılımin normal olmadığı görülmektedir.

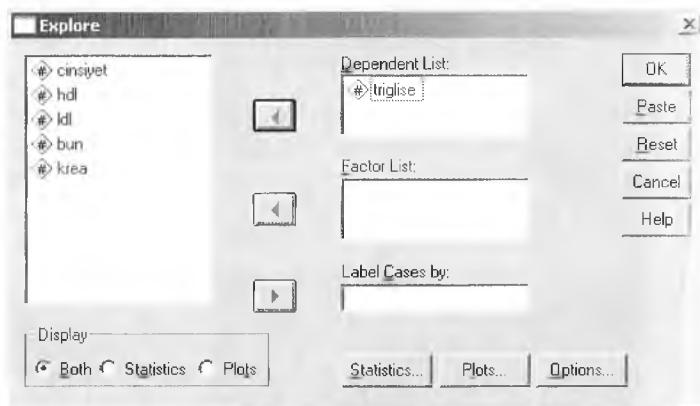


Şekil 8
Örnek 34'teki çalışma grubunda Ca değerlerinin histogramı ve beklenen teorik normal eğrisi.

Explore menüsü

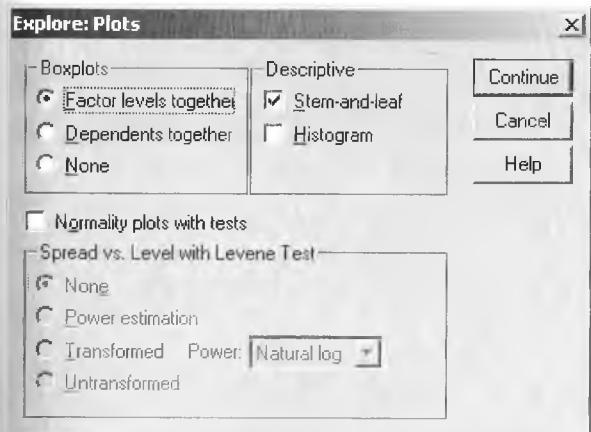
"Analyze >> Descriptive Statistics >> Explore" diyalog kutusundan normal dağılımla ilgili testler ve grafiklerin elde edilebildiği Bölüm 3'te anlatılmıştı. Yine bu kutudaki *Plots* seçeneği ile açılan penceredeki *Box-plots* ve *Descriptives* kutularından histogram, gövde-yaprak grafiği ve boksör torbası grafiği de çizdirilebilir.

Grafiği çizdirilecek değişkenin *Dependent List*'e alınması gereklidir. Eğer tüm grubun grafiği çizdirilecekse, *Factor List*'te hiç bir değişkenin bulunmaması gereklidir. Altgrupların grafikleri çizdirilecekse, altgrupları belirleyen değişken *Factor List*'e alınmalıdır (Şekil 9).

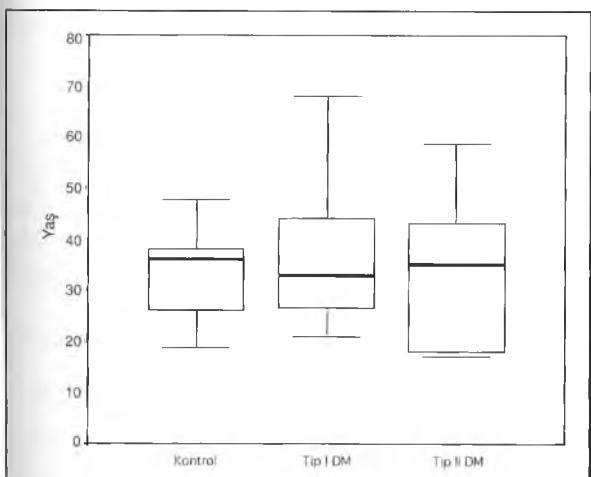


Şekil 9
SPSS'te "Explore"
diyalog kutusu.

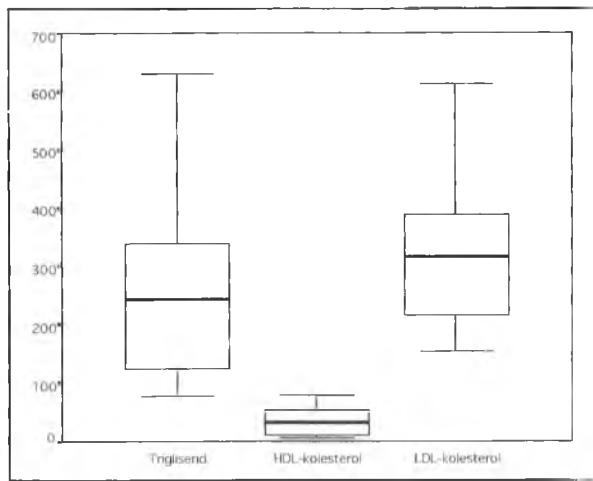
Plots diyalog kutusunu açığınızda tanımlayıcı grafiklerle ilgili seçeneklerin yer aldığı göreceksiniz. *Descriptive* kutusunda gövde-yaprak grafiği önceden seçilmiş durumdadır. Bu kutudan histogramı da seçebilirsiniz. *Boxplots* kutusunda ise *Factor levels together* önceden seçilmiş durumdadır (Şekil 10). Tip I DM, tip II DM ve kontrol grubunun yaşlarının bu şekilde çizilmiş boksör torbası grafiği Şekil 11'deki gibi olacaktır. Tüm grubun yaş, trigliserid ve kreatinin değerlerine ilişkin boksör torbası grafiği ise *Dependents together* seçimi ile çizildiğinde Şekil 12'deki gibi olacaktır.



Şekil 10
SPSS'te "Explore: Plots"
diyaloğunu.



Şekil 11
"Factor levels
together" seçimi ile çizilmiş
boksör torbası grafiği.



Şekil 12
"Dependents together"
seçimi ile çizilmiş boksör
torbası grafiği.

bölüm 8

ORTALAMALARIN KARŞILAŞTIRILMASI

Ortalama, oran, ortanca

Ortalamalar farklı mı?

Oran mı, ortanca mı?

Ortalama mı, ortanca mı?

Güven aralığı mı, hipotez testi mi?

Bağımsızlık ve bağımlılık nedir?

Grupların bağımsızlığı

Değişkenlerin bağımsızlığı

Ortalama ya da ortancaların karşılaştırılabilen çalışma düzenleri

Bölüm 9, 10 ve 11'de grup ortalamalarının ve ortancalarının karşılaştırılması ile ilgili istatistik analiz yöntemleri anlatılacaktır. İstatistik analiz yöntemlerine giriş yaptığımız bu bölüme özgü olarak, bazı formüller ve matematik işlemlere yer verilecektir. Bunun amacı yalnızca istatistik analizin mekaniginin daha iyi anlaşılabilmesini sağlamaktır. Zaten bu kitabın amacı, teorik istatistik bilgileri aktarmak değil, biyoistatistik bilimi ile ilgili temel kavramları, hatta aşırı sadeleşmenin getirdiği teorik eksikliklerin de farkında olarak anlatmaya çalışmaktadır. Eğer istatistik analizlerinizi elle yapacaksanız, bol formül içeren çok sayıda kitap vardır. Eğer bilgisayarla istatistik analiz yapacaksanız, zaten formülleri bilmenize gerek yoktur. Ama istatistik analiz doğru yöntem seçimi ve analiz sonuçlarının doğru yorumlanmasıdır. İşte bu iki nedenle, istatistik analizinin mekanigini az da olsa bilmekte yarar vardır.

ORTALAMA, ORAN, ORTANCA

Ortalamalar farklı mı?

Araştırma yapmanın amaçlarından biri, belki de en önemlisi çeşitli gruplar arasında fark olup olmadığını saptanmasıdır. Serumda, plevral effüzyonda, eklem sıvısında vb. düzeyi ölçülen bir değişkenin hastalık grupları arasında farklı olduğunun saptanması, tanışal açıdan büyük önem taşır. Ya da bir ilaçla verilen yanıtın ya da görülen etkilerin erkek ve kadınlarda farklı ol-

ması tedavi seçiminde hastanın cinsiyetini göz önüne almak gibi bir yaklaşımı gündeme getirebilir.

ÖRNEK 35

Yeni bir antiinflamatuar ilacın romatoid artritte ve seronegatif artritte etkinliğinin araştırıldığı bir çalışmada aktif romatoid artritli ve aktif sero-negatif artritli 50'şer hastada iki ay süre ile tedavi uygulandıktan sonra, hastaların tedaviye yanıtları klinik olarak değerlendiriliyor. Ayrıca hastalık aktivitesini gösteren bazı değişkenler de (**eritrosit sedimentasyon hızı gibi**) ölçülüyor.

Gruplara özgü değişkenlik gösteren ölçütlerle çalışırken, öncelikle karşılaştırılan özelliği tanımlamak gereklidir. Örnek 35'te grupları temsil eden ölçüt olarak her grupta tedaviye yanıt veren ve vermeyen hastaların oranları karşılaştırılabilcegi gibi, hastalık aktivitesi ile ilişkili bir değişkenin, örneğin eritrosit sedimentasyon hızının değişiminin de gruplar arasında farklı olup olmadığı asıl ilgi konusu olabilir. Birinci durumda "yanıt tipi" olarak adlandırabileceğimiz değişken, eğer "yanıt var" ve "yanıt yok" gibi iki değer alabiliyorsa, bu değişkene ikili nominal değişken denildiğini biliyorsunuz. Bu durumda gruplara özgü ölçüt, orandır. Bu nedenle istatistik analizde yapılması gereken romatoid artrit ve sero-negatif artrit gruplarında "yanıt verenlerin gruptaki toplam hasta sayısına oranları"nın karşılaştırılmasıdır. Grup oranlarının karşılaştırılması ile ilgili analiz yöntemleri ayrıntılı olarak Bölüm 12'den başlayarak anlatılmıştır.

Oran mı, ortanca mı?

Eğer "yanıt tipi" değişkenine iki değer yerine "yanıt yok", "minimal yanıt var", "orta derecede yanıt var", "iyi derecede yanıt var", "çok iyi derecede yanıt var" ve "tam yanıt var" gibi ikiden çok ve sıralı olarak da değerler verilmişse, bu kez bu değişkenin nominal değil ordinal değişken olarak adlandırılmasını da biliyorsunuz. Bu durumda, bir önceki paragrafta olduğu gibi yine grup oranları karşılaştırılabilir. Öte yandan ordinal değişkenlerin merkezi eğilim ölçütünün ortanca olduğunu hatırlarsanız, grup ortancalarının da karşılaştırılması düşünülebilir. "Yanıt tipi" değişkenini "yanıt yok"tan "tam yanıt"a doğru 0'dan 5'e kadar rakamlarla belirtip, bir "yanıt skoru" oluşturursanız, iki grup arasında "yanıt skoru"nun ortancalarının farklı olup olmadığını inceleyebilirsiniz. Yalnız ordinal değişkenlerin ortancalarını karşılaştırırken, bir iki noktaya dikkat etmek gereklidir.

ÖRNEK 36

Meme kanserinde depresyon ile ilgili bir çalışmada evre I-IV olarak dört evrede sınıflandırılan hastalarda tedaviden önce ve sonra 1'den 100'e kadar değer alabilecek bir skalada "depresyon skorları" ölçülüyor. Tümör evresi ve depresyon skorlarında tedavi öncesi ve sonrası fark olup olmadığı inceleniyor.

Örnek 36'da 1 ile 100 arasında çeşitli değerler alabilecek "depresyon skoru"nın ortanca değerinin, çalışma grubunun merkezi eğilimini temsil ettiği rastathıkla söyleylenebilir. Örneğin "depresyon skoru"nın tedavi öncesi ortanca değerinin 55 olarak ve 25 persentil ve 75 persentil değerlerinin de sırasıyla 23 ve 68 olarak bildirilmesi "depresyon skoru"nın dağılımı hakkında bir fikir verir.

Öte yandan yalnızca dört değer alabilecek "tümör evresi"nden söz ederken, "ortanca tümör evresi III ve 25 persentil ve 75 persentil değerleri I ve IV'tür" cümlesiindeki bilgilerin çok anlamlı olmadığı anlaşılabılır. Ortanca tümör evresi yerine hastaların %12'sinin evre I, %34'ünün evre II, %36'sının evre III ve %18'inin evre IV'te olduğunu bilinmesi daha bilgi vericidir.

Gruplar karşılaştırılırken, bir ordinal değişkenin ortancasının mı, yoksa değişken düzeylerinin oranlarının mı ele alınmasının daha doğru olduğu ile ilgili seçim yaparken, değişkenin düzey sayısının oldukça önemlidir.

Bu örnekte 100 farklı düzeyi olan bir ordinal değişken olan "depresyon skoru"nu incelerken ortancaların, dört farklı düzeyi olan bir ordinal değişken olan "tümör evresi"ni incelerken oranların kullanılmasının daha doğru olduğu açık olarak görülmektedir. Bu seçimi yaparken, değişkenin düzey sayısının kaçın altındaysa oran, kaçın üstündeyse ortancanın tercih edileceği konusunda belirli bir rakam yoktur. Tıp literatüründe Apgar skoru gibi 10 düzeyli bir ordinal değişken için düzey sayısının yeterli görülmüş ortanca kullanılmaktadır. Önemli olan ortanca değerin grubun merkezi eğilimini iyi temsil ettiği konusundaki yorumdur.

Ortalama mı, ortanca mı?

Örnek 35'e geri dönersek, eritrosit sedimentasyon hızı gibi sayısal değişkenler söz konusu olduğunda, değişkenin ortalamasının mı, ortancasının mı grubun merkezi eğilim ölçütü olarak kullanılmasının daha doğru olduğu incelenmelidir. Bölüm 5'te merkezi eğilim ölçütü olarak en sık kullanılan iki ölçüt tipinin ortalama ve ortanca olduğunu görmüş ve hangi durumlarda hafifinin kullanılmasının daha doğru olduğunu ilişkin bazı noktalardan söz etmişlik. İşte gruplardaki sayısal değişkenleri karşılaştırırken de, aynı kriterler geçerlidir. Simetrik ve normal dağılan veriler için ortalama daha uygundur.

Dağılım eğrisi simetrik olmayan veriler için ise ortancalar ele alınmalıdır.

Güven aralığı mı, hipotez testi mi?

Bölüm 6'da güven aralıklarının karşılaştırılması ve hipotez testlerinin istatistik analizde kullanılan iki yöntem olduğundan ve elde edilecek sonuç açısından ikisi arasında hiç bir fark bulunmadığından söz edilmiştir. Ortalamaların karşılaştırılmasında da bu geçerlidir. Güven aralıklarının kullanılması ile bir çalışma grubunda gözlenen ortalamanın evrene ekstrapole edilirken, ne kadar değişim göstereceği daha anlaşılır bir şekilde ortaya konulurken, alışkin olduğumuz p değerlerinin hesaplanması hipotez testlerinin kullanılması ile olur.

Burada tekrar vurgulamakta yarar var. Herhangi bir istatistik analiz sırasında hipotez testi ya da güven aralıklarının kullanılması arasında istatistiksel anlamlılık elde etme açısından hiç bir fark yoktur. İkişi birbirinin yerine kullanılabilir.

BAĞIMSIZLIK VE BAĞIMLILIK NEDİR?

Grupların bağımsızlığı

Bağımsız terimi istatistikte gruplar ve değişkenleri tanımlarken farklı anlamlarda kullanılır. Bağımsız gruplar denince, her grupta farklı deneklerin yer aldığı, bir gruptaki bir denegin aynı zamanda başka gruptarda da bulunmalığı anlaşıılır.

Aşında grplarda bulunan kişiler değil, kişilere ait verilerdir. Bu nedenle bazen aynı kişilerin farklı verileri farklı grplarda yer alabilir, ama "denek" teriminden kişinin kendisi değil, kişiye ait veri kastedildiği için, grplar yine de bağımsız olabilir.

ÖRNEK 37

Nötropenik ateş epizodlarının tedavisinde amikasin+seftadizim ile tek başına imipenem tedavisinin etkinliği karşılaştırılıyor. Yüz hastanın 50'sindeki nötropenik ateş epizodunda birinci tedavi, 50'sindeki nötropenik ateş epizodunda ikinci tedavi uygulanması planlanıyor.

Örnek 37'deki çalışmada her hasta ya birinci grupta ya da ikinci grupta yer alacağı için grplar bağımsız olacaktır. Ama bir hastada farklı zamanlarda birden çok ateş epizodu olabileceği düşünülürse, çalışmadaiki iki grubu 70 hastanın 100 epizodu da oluşturabilir. Bu durumda birden çok epizodu olan hastaların bazlarına farklı epizodlarında farklı tedaviler verilmiş olabileceği için,

aynı hastanın farklı epizodları farklı gruplarda yer alabilir. Bu durumda da gruplar bağımsız olarak düşünülebilir. Örneğin tedavi başlandığındaki beyaz küre sayılarının ortalamaları karşılaştırıldığını düşünelim. Farklı zamanlarda iki tedavinin de verildiği bir hastaya amikasin+seftazidim verildiği zamanda ki beyaz küre sayısı ile imipenem verildiği zamandaki beyaz küre sayısı bağımsızdır. O halde bu hastanın iki farklı epizoduna ilişkin bu verilerin iki farklı grupta yer alması grupların bağımsızlığını bozmaz.

Her ne kadar farklı gruplarda kısmen de olsa aynı deneklere ait verilerin bulunması, mutlaka grupların bağımsız olmadığı anlamına gelmese de, yine de mümkün olduğu kadar bundan kaçınmak gereklidir. Özellikle insan, hayvan gibi biyolojik deneklerle çalışırken, kimi deneklerin uyarılara yanıtının her uyarıyla karşılaşlıklarını zaman aynı olabileceği düşünülürse, bu şekildeki çalışma düzenlerinin sakıncaları daha iyi anlaşılabilir. Örneğin Örnek 37'deki çalışmada, seftazidime idyosenkrazik aşırı duyarlılık yanıtını verdiği bilinen bir hastanın tesadüfen üç kez amikasin+seftazidim grubuna aldığı düşündür. Bu durumda bu tedavi protokolünün %6 (3/50) gibi gerçekte olduğundan daha yüksek bir yan etki oranı görülecektir.

ÖRNEK 38

Elli tip I DM hastasında hiperglisemik ve normoglisemik dönemde fruktozamin düzeyleri ölçülüyor. "Hiperglisemik dönemdeki ortalama fruktozamin değeri, normoglisemik dönemdeki ortalama fruktozamin değerinden farklı mıdır?" sorusuna yanıt ararken bu iki grubun bağımsız olduğu söylemenemez. Çünkü her iki grupta da aynı hastaların fruktozamin değerleri yer almaktadır. Daha önemlisi her hastanın iki farklı dönemdeki değerleri, birbirlerinden bağımsız değildir. Örneğin bir hastanın hiperglisemik dönemde 6.0 olan fruktozamin değeri normoglisemik dönemde 4.5'a düşerken, başka bir hastanın fruktozamin değeri 4.0'dan 2.5'a düşebilir.

Değişkenlerin bağımsızlığı

Değişkenler arasında ilişki olup olmadığını araştırırken, değişkenlerden bazılarına bakarak, diğer değişkenlerin ne derece değişeceğini kestirilmeye çalışılır.

Lösemide IL-3 düzeyinin normal değerlerden daha yüksek olması yanında, "hastaların beyaz küre sayıları artınca, IL-3 düzeyi de artıyor mu, yoksa beyaz küre sayısı ile IL-3 düzeyi arasında bir ilişki ya da bağıntı yok mu?" soruları beyaz küre sayısı ile IL-3 düzeyi arasında ilişki olup olmadığını anlamaya yönelikdir. Bu örnekte beyaz küre sayısı bağımsız, IL-3 düzeyi bağımlı değişkendir. Çünkü IL-3 düzeyinin, beyaz küre sayısıyla ne kadar ilişkili ol-

duğu, ya da beyaz küre sayısına ne kadar "bağımlı" olduğu araştırılmaktadır.

Aynı şekilde "DM'un iki farklı tipi arasında fruktozamin düzeyleri açısından fark var mıdır?" sorusu ile merak edilen şey, fruktozamin düzeyinin DM'un tipine "bağımlı" olup olmadığıdır. Burada "DM'un tipi" bağımsız, "fruktozamin düzeyi" bağımlı değişkendir. Bağımsız değişken, bilgisayar programlarında "factor", "independent variable", "grouping variable", "control variable" gibi adlarla da anılmaktadır.

Grup ortalamaları (ya da ortancaları) karşılaştırılırken, aslında yapılan nominal bir değişkenle (grupları belirleyen değişken), sayısal bir değişken (karşılaştırılacak ortalamaları belirleyen değişken) arasında ilişki olup olmadığıının araştırılmasıdır.

ORTALAMA YA DA ORTANCALARIN KARŞILAŞTIRILABILECEĞİ ÇALIŞMA DÜZENLERİ

ÖRNEK 39

Lösemide interlökin-3 (IL-3) düzeyinin normalden farklı olup olmadığıın araştırılmak isteniyor. Bu amaca uygun çeşitli çalışma düzenleri uygulanabilir:

- a. Lösemili bir grup hastada IL-3 düzeyi bakılır. Ortalama IL-3 düzeyinin normal değerlerinden farklı olup olmadığı test edilebilir.
- b. Lösemili hastalardan oluşan bir grup ve normal deneklerden oluşan bir grupta IL-3 düzeyi bakılır. İki grubun ortalama IL-3 düzeyi birbirleriyle karşılaştırılır.
- c. Lösemili bir grup hastada tedaviye başlamadan önce IL-3 düzeyi bakılır. Tedavi sonrası remisyona girdikten sonra tekrar bakılan IL-3 düzeylerinin ortalaması, tedavi öncesi ortalama IL-3 düzeyi ile karşılaştırılır.
- d. Lösemili, lenfomali ve normal deneklerden oluşan üç grup hastada IL-3 düzeyi bakılır. Üç grubun IL-3 düzeyi ortalaması karşılaştırılır.
- e. Lösemili bir grup hastada tedaviden önce, tedavinin birinci, ikinci ve dördüncü haftalarında, remisyona girdikten sonra ve relaps olduktan sonra IL-3 düzeyi bakılır. Farklı dönemlerde bakılan IL-3 düzeyi ortalamaları karşılaştırılır.
- f. Lösemili ve lenfomali iki grup hastada tedaviden önce, tedavinin birinci, ikinci ve dördüncü haftalarında, remisyona girdikten sonra ve relaps olduktan sonra IL-3 düzeyi bakılır. Farklı dönemlerde bakılan IL-3 düzeyi ortalamaları karşılaştırılır.

Örnek 39'da IL-3 düzeyine ilişkin verilerin normal dağıldığını varsayırsak, her altı çalışma düzeneinde de grup ortalamalarının karşılaştırılmasına ilişkin analiz yöntemlerinin kullanılması gerektiği anlaşıılır. Bu çalışma düzenlerini özetlersek;

- a. Tek grup içeren çalışma düzenei
- b. İki grup içeren çalışma düzenei
- c. Bir grupta iki farklı zamanda ölçüm yapılan çalışma düzenei
- d. İkiden çok grup içeren çalışma düzenei
- e. Bir grupta ikiden çok zamanda ölçüm yapılan çalışma düzenei
- f. İki ya da daha çok grupta iki ya da daha çok zamanda ölçüm yapılan çalışma düzenei

TEK GRUP ORTALAMASI

bölüm 9

Evren ortalamasının güven sınırları

Değişkenliğin etkisi nasıldır?

Denek sayısının etkisi nasıldır?

Hipotez testi (tek grup ortalaması için t testi)

Verilerin dağılımı normal değilse

Dönüşümler

SPSS'te tek grup ortalaması testi nasıl yapılır?

T testi

Bir grubun ortalamasının bilinen bir değerden farklı olup olmadığını belirleyen üç etken vardır. Bunlar (1) grup ortalamasının karşılaştırılan değerden ne kadar yüksek ya da düşük olduğu, (2) hasta değerlerinin ne kadar değişim gösterdiği ve (3) örneklem büyütüğü, yani hasta sayısıdır. Ortalama, karşılaştırılan değerden ne kadar farklısa, aradaki farkın önemli olma olasılığı o kadar fazladır. Ancak eğer hasta değerleri büyülü küçülü ise, yani değişkenlik çok fazla ise, olasılık azalır. Aynı şekilde hasta sayısı da azsa, farkın önemli olması olasılığı azalır.

ÖRNEK 40

İnterlökin-III (IL-3) düzeyinin normal bireylerde 10.0 pg/mL olduğu bilinmektedir. Lösemide IL-3 düzeyinde değişme olup olmadığını araştırmak amacıyla 49 lösemik hastanın bulunduğu bir çalışma yapılmıştır. IL-3 değerlerinin ortalaması 12.3 pg/mL ve standart sapması 3.4 pg/mL olarak saptanmıştır.

Burada yanıtlanacak soru “lösemi grubu IL-3 açısından normal değerlerin elde edildiği grupta aynı evrenden mi seçilmişdir?” ya da daha açık bir tanımlamayla “12.3, 10.0’dan farklı mıdır?” dır. Eğer verilerin dağılımı normalse bu soru yanıtlanabilir.

Denek sayısı 30'un üzerinde ise -dağılımin normalden çok fazla saptalar göstermemesi koşulu ile-, dağılımin normal olduğu varsayılabılır.

Tek grup ortalamasının bilinen bir değerden farklı olup olmadığı sorusunu yanıtlamak için iki yol vardır: (1) evren ortalamasının güven aralığının belirlenmesi ya da (2) tek grup ortalaması için hipotez testi.

EVREN ORTALAMASININ GÜVEN SINIRLARI

Evrenin standart sapması her zaman bilinmeyebilir. Örneğin serum kalsiyum düzeyi gibi uzun süredir bilinen değerler için ortalama ve standart sapma yerine alt ve üst sınırlar kullanılır. Öte yandan IL-3 düzeyi gibi süredir üzerinde çalışılan değişkenlerin normal düzeyleri genellikle ortalama ve standart sapması ile birlikte verilir. Evren değerlerinin standart sapmasının bilindiği ve bilinmediği durumlarda farklı formüller uygulanır. Genellikle evren standart sapmasının bilinmemesi nedeniyle yalnız bu durum için örnek verilmiştir.

Evren standart sapması bilinmiyorsa;

$$\%95 \text{ güven sınırları} = \text{Ortalama} \pm (t_{n-1, 0.05})(\text{SEM})$$

$$\text{Ortalamanın standart hatası (SEM)} = \frac{\text{SD}}{\sqrt{n}}$$

$$n = \text{Denek sayısı}$$

Bölüm 6'daki tartışmalardan, hipotez testleri ile hesaplanan ve test istatistik adı verilen rakamların bazı tablolardaki değerlerle karşılaştırılması sonucunda p değerinin elde edildiğini hatırlayacaksınız.

Tablolardaki değerlerden hangisinin hesaplanan değerle karşılaşılacağı seferken α yanılma düzeyi ve serbestlik derecesinin belirleyici olduğundan ve serbestlik derecesinin her hipotez testi için farklı olduğundan söz edilmiştir. Tek grup ortalaması ile ilgili güven sınırlarının hesaplanması ve hipotez testinde serbestlik derecesi denek sayısından bir eksik olarak tanımlanmıştır.

Bu formüldeki $t_{n-1, 0.05}$ değeri, t tablosundaki iki yönlü α yanılma düzeyi 0.05 ve serbestlik derecesi 48'e (yani $n-1$) karşılık gelen t değeridir. Bu değer yaklaşık 2.01'dir. Formülde yerine koyarsak, lösemi grubundaki IL-3 düzeyinin %95 güven sınırları $12.3 \pm (2.01)(3.4/7) = 12.3 \pm 0.98$, yani 11.3 ve 13.3'dür (Güven sınırlarının 11.32 ve 13.28 olarak verilmemişine dikkat ediniz). Bunun nedeni ortalama IL-3 düzeyinin 12.30 olarak değil, 12.3 olarak verilmesidir. Yani noktanın sağında önemli rakam sayısı "bir"dir. Öte yandan 0.98'de noktanın sağındaki önemli rakam sayısı "iki"dir. Bu nedenle 0.98 rakamı, 12.3'teki gibi

noktanın sağındaki rakam sayısı “bir” olacak şekilde 1.0’ a yuvarlanmıştır). Yanılma düzeyi 0.05 seçildiği için lösemili hastaların seçildiği evrenin IL-3 ortalamasının %95 olasılıkla 11.3 ile 13.3 arasında olduğu anlaşılmaktadır. Normal değer olan 10.0 bu güven sınırlarının dışında olduğu için “lösemili hastalar evreninin IL-3 ortalaması %95 olasılıkla 10.0’ dan farklıdır” denebilir.

Değişkenliğin etkisi nasıldır?

Hasta değerlerindeki değişkenliğin etkisini incelemek için standart sapmanın 3.4 yerine 10.3 olduğu, yani değerlerin değişkenliğinin arttığı durumda aynı formülü uygulayalım.

Bu durumda %95 güven sınırları $12.3 \pm (2.01) (10.3/7) = 12.3 \pm 2.96$, yani 9.3 ve 15.3 olacaktır. Normal değer olan 10.0, lösemili hastaların IL-3 ortalamasının evrende dağılım gösterdiği bu güven aralığına dahil olacağı için “lösemili hastaların IL-3 ortalaması normal değerden farklıdır” yorumunun yanılma olasılığı %5’ten büyük olacaktır. Bu nedenle yokluk hipotezi, yani fark olmadığı kabul edilmelidir.

Denek sayısının etkisi nasıldır?

Bu kez de hasta sayısının etkilerini incelemek için, çalışma grubundaki hasta sayısının 49 yerine 16 olması durumunda aynı formülü uygulayalım.

Hasta sayısının 16 olduğu durumda formüle konulacak t değeri olarak Bölüm 6 Tablo 2’de 15 serbestlik derecesine karşılık gelen 2.13 konulmalıdır. %95 güven sınırları $12.3 \pm (2.13) (3.4/4) = 12.3 \pm 1.81$, yani 10.5 ve 14.1 olacaktır. Normal değer olan 10.0 güven aralığının dışında olduğu için, “lösemi grubunun IL-3 ortalaması, %95 olasılıkla normal değerden farklıdır” denebilir.

HİPOTEZ TESTİ (TEK GRUP ORTALAMASI İÇİN T TESTİ)

Güven sınırlarının hesaplanması baştan belirlediğimiz α yanılma düzeyine karşılık gelen t değerini tablodan bulup, formülde yerine koymuştık. Hipotez testinde ise, işlem tersinden yapılır. Önce t değeri hesaplanır; daha sonra bu değere karşılık gelen p değerinin α yanılma düzeyi ile karşılaştırılması yapılır. Hesaplanan t değerine tabloda karşılık gelen α değeri, yokluk hipotezinin reddedilmesi, yani grubun ortalamasının normalden farklı olduğunu kabul edilmesi ile ne kadar yanlışlığımızı belirtir. Eğer hesaplanan t değerinin mutlak değeri, tabloda α yanılma düzeyi 0.05 ve serbestlik derecesi n-1'e karşılık gelen t değerinden büyük ise yokluk hipotezinin reddedilmesi ile yanılma olasılığı %5’ten küçük demektir.

Tek grup ortalaması için hipotez testinde t değeri aşağıdaki formülle hesaplanır.

$$t_{n-1, 0.05} = \frac{\text{Örneklem ortalaması - normal değer}}{\text{SEM}}$$

$$\text{Ortalamanın standart hatası (SEM)} = \frac{\text{SD}}{\sqrt{n}}$$

$$n = \text{Denek sayısı}$$

Örnek 40'daki çalışma için hipotez testi uygularsak,

H_0 hipotezi = "Lösemi grubunun IL-3 ortalaması normal değerden farklı değildir"

H_1 hipotezi = "Lösemi grubunun IL-3 ortalaması normal değerden farklıdır"

$$t = \frac{12.3 - 10.0}{3.4 / \sqrt{7}} = \frac{2.3}{0.49} = 4.694$$

$$\text{Serbestlik derecesi} = 49 - 1 = 48$$

Tabloda 48 serbestlik derecesine ve 0.05 yanılma düzeyine karşılık gelen t değeri 2.01'dir. Hesaplanan t değeri, bu değerden büyük olduğu için %5 yanılma olasılığı ile H_0 hipotezi reddedilebilir, yani "en az %95 olasılıkla lösemi grubunun IL-3 ortalaması normalden farklıdır" denebilir. Ya da daha doğru bir tanımlama ile "Aynı büyülüklükte bir örneklemle yapılacak 100 çalışmanın en az 95'inde lösemi grubun IL-3 ortalaması normalden farklı olacaktr" denebilir. Hatta hesaplanan t değeri 0.001 yanılma düzeyine karşılık gelen tablo değeri olan 3.5'tan bile büyük olduğu için "en az %99.99 olasılıkla lösemi grubun IL-3 ortalaması normalden farklıdır" denebilir.

Güven aralığı yöntemi ile hipotez testi arasında sonuç açısından fark yoktur. İkişinde de ortalamanın farklı olup olmadığı konusunda kesinlikle aynı sonuca ulaşılır. Bu nedenle hangi yöntemin kullanılacağı tamamen bir tercih sorunudur. Son yıllarda tip literatüründe hipotez testi yanında ya da yalnız güven sınırlarının belirtilmesi yönünde bir eğilim vardır.

Hesaplanan t değeri 0'dan küçük de olabilir. Örneğin lösemi grubunda IL-3 düzeyi 10.0, normal IL-3 düzeyi 12.3 pg / mL olsaydı, t değeri -4.694 olarak hesaplanacaktı. Hesaplanan t değerinin mutlak değeri, tablodaki t değeri ile karşılaştırıldığı için bulunacak yanılma olasılığı değişmez.

Verilerin dağılımı normal değilse

Yukarıdaki örneklerde IL-3 değerlerinin normal dağılım gösterdiğini varsayıdık, çünkü tek grup ortalamasının normal değerler karşılaştırılmasındaki ana varsayımdı budur. Hasta sayısı 30'dan fazlaysa genellikle sorun olmaz, ama daha az sayıda hasta varsa, değerlerin dağılıminın normal olup olmadığı incelenmelidir. Normal dağılım göstermeyen değerlerin ortalamasının t testi ile doğrudan test edilmesi ile gerçekte olduğundan daha küçük bir p değeri, ya da daha dar güven aralığı hesaplanır. Bu durumda grubun ortalaması normalden farklı olmadığı halde, farklılaşmış gibi sonuç elde edilebilir. Yani istatistik analizin tip I hatası artar.

Verilerin normal dağılmadığı durumlarda genel olarak iki çözüm vardır:

1. Verilere dönüşüm (*transformation*) uygulayarak, dönüştürülmüş verilerin normal dağılıma uymasının sağlanması ve bu verilere test uygulanması.
2. Verilerin dağılımı ne olursa olsun uygulanabilen eşdeğer testlerin uygulanması.

Parametrik testlerde sağlanması gereken koşulların başında verilerin normal dağılıma uyması gelir. Verilerin normal dağılım göstermediği, denek sayısının az olduğu ya da diğer parametrik test uygulayabilme koşullarından herhangi birinin sağlanamadığı durumlarda, parametrik testlerin eşdeğeri olan nonparametrik testler uygulanabilir. Tek grup ortalaması için eşdeğer nonparametrik bir test yoktur. Bu nedenle normal dağılım göstermeyen verilere dönüşüm uygulanmalıdır.

DÖNÜŞÜMLER

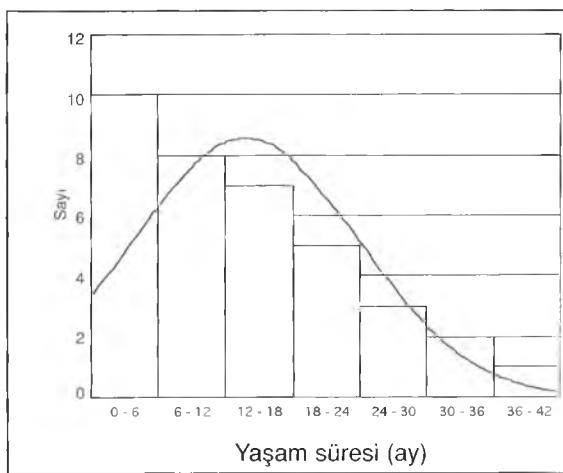
Normal dağılımda değişken değeri ortalamaya yakın olan denek sayısının fazla olduğu ve ortalamadan daha küçük ya da büyük değerlere doğru gidilmesi, denek sayılarının azaldığını biliyoruz.

Normal dağılım eğrisi çizildiğinde, çan şekline benzer. Bir gruptaki değerlerin çögünün ortalamadan küçük olduğu, yalnızca bir kaç değerin ortalamadan büyük olduğu durumda dağılım eğrisi çok farklı olacaktır. Örneğin 100 Lanser hastasının tamdan sonraki yaşam sürelerine bakıldığında, hastaların %50'unun yaşam süresinin kısa olduğu ve süre uzadıkça, yaşayan hasta sayısının azaldığı görülecektir (Şekil 1).

Yaşam süresinin histogramı çizildiğinde soldaki (başlangıçta) çubuğun tepe noktasıının en üst noktada olduğu ve sağa doğru gittikçe tepe noktalarının sağa doğru indiği izlenecektir. Bunun gibi normal dışı dağılım gösteren verilere ilişkin ortalamaları test ederken, verileri olduğu gibi kullanmak hatalı olacaktır. Bu durumlarda değerlerin logaritması, karekökü, vb. matematik bir işlem uygulandıktan sonraki değerlerin dağılımı normal dağılıma yaklaşır.

En sık kullanılan dönüşüm yöntemlerinden biri de sıra dönüşümüdür (*rank transformation*).

Sıra dönüşümünde değişken değerleri küçükten büyüğe doğru dizildikten sonra, en küçük değere 1, ikinci değere 2, ... gibi sıra numaraları verilir ve testler sırasında verilerin gerçek değeri yerine bu sıra numaraları kullanılır. Özellikle bir taraftan basık dağılım gösteren değişkenler için sıra dönüşümü oldukça uygundur.



Şekil 1

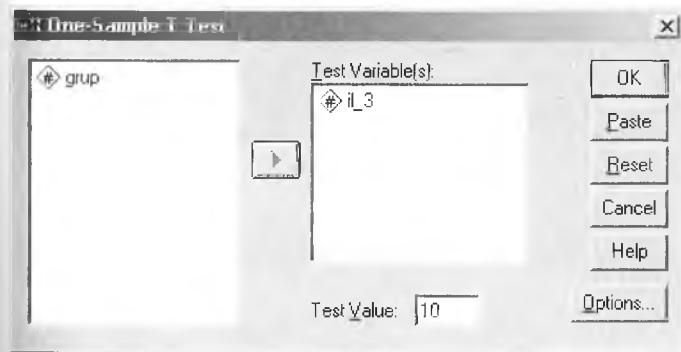
Kanser hastalarında yaşam süresinin histogramı.

SPSS'TE TEK GRUP ORTALAMASI TESTİ NASIL YAPILIR?

T testi

Statistics >> Compare Means >> One-Sample T Test

Yukarıdaki seçimle **One-Sample T Test** diyalog kutusu açılır (Şekil 2). **Test** edilecek değişken *Test variable(s)* kutusuna seçilmelidir. *Test value* kutusunun grup ortalamasının karşılaşacağı değerin yazılması gereklidir. Örnek 40'da grup IL-3 ortalamasının 10 pg / mL'den farklı olup olmadığı test edileceğse, *Test value* kutusuna '10' yazılmalıdır. *Options* penceresinden önceden %95 olarak seçilmiş olan güven aralığını değiştirmeye olağanı vardır.



Şekil 2
SPSS'te "One-Sample T Test" diyalog kutusu.

Sonuçta Şekil 3'deki tablo elde edilir. *One-Sample Statistics* tablosunda grubun ortalama, standart sapma (SD) ve standart hata değerleri verilmiştir. *One-Sample Test* tablosunda ise t değeri, serbestlik derecesi (*df*), p değeri (*Sig. 2-tailed*) ve %95 güven sınırları verilmiştir. P değeri 0.000, yani <0.0001 'tür. Grup ortalaması ile '10' arasındaki farkın %95 olasılıkla 1.328 ve 3.285 arasında olduğu görülmektedir. Güven aralığı 0'ı (sıfır) içermediği için %95 olasılıkla grup ortalamasının 10'dan farklı olduğu anlaşılmaktadır.

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
IL_3	49	12.31	3.41	.49

One-Sample Test						
	Test Value = 10					95% Confidence Interval of the Difference
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
IL_3	4.738	48	.000	2.31	1.33	3.29

Şekil 3 Örnek 40'daki çalışmada lösemide IL-3 ortalamasının '10' ile karşılaştırıldığı t testi sonuçları.

bölüm 10

İKİ ORTALAMANIN KARŞILAŞTIRILMASI

Bağımlı (eş) iki grup ortalamasının karşılaştırılması

Güven sınırları

Hipotez testi (bağımlı iki grup ortalaması için Student t testi)

Verilerin dağılımı normal değilse

Bağımsız iki grup ortalamasının karşılaştırılması

Güven sınırları

Hipotez testi (bağımsız iki grup ortalaması için Student t testi)

Grup varyanslarının eşitliği

Student t testi varsayımları sağlanamıyorsa

SPSS'te bağımlı iki grup ortalaması nasıl karşılaştırılır?

Student t testi

Wilcoxon işaret testi

SPSS'te bağımsız iki grup ortalaması nasıl karşılaştırılır?

Student t testi

Mann-Whitney U testi

BAĞIMLI (EŞ) İKİ GRUP ORTALAMASININ KARŞILAŞTIRILMASI

Klinik çalışmaların çoğu birden çok grup bulunur. Örneğin bir ilaçın etkinliğini araştırmak amacıyla yapılan bir çalışmaya alınan iki gruptan birine bu ilaç verilirken, diğerine başka bir ilaç ya da placebo verilerek, iki grup birbirile karşılaştırılır. Ya da bir hastalıkta kanda bir maddenin düzeyinin normalden farklı olup olmadığını araştırılması için, sağlıklı deneklerde de aynı maddenin kandaki düzeyleri ölçülüp iki grubun ortalamalarılaştırılır.

Bu iki örnekte de gruplar bağımsızdır, yani çalışmaya alınan denekler ya bir gruptadır, ya da öteki gruptadır. Her iki grupta da yer alan bir denek bulunmaz. Bu çalışma döneminde grupların yaş, hastalığın şiddeti, vb. gibi birçok özellikler açısından benzer olmaları sağlanmalıdır.

Bağımsız gruplarda her türlü çabaya karşın bazı özellikler açısından farklılıklar olabilir. Özellikle insan, hayvan gibi biyolojik deneklerde boy, vücut ağırlığı, yaş vb. bir çok özellikler açısından varyasyon çok belirgindir.

Bağımsız gruplarla çalışmanın bu sakıncaları nedeniyle, her denegenin her iki

grupta da yer aldığı klinik çalışma düzenleri de sıkılıkla kullanılmaktadır. "Önce-sonra", ya da "pretest-posttest" tipindeki çalışmalar bu tip çalışmaların prototipidir. Örneğin hipolipidemik bir ilacın etkinliğini denemek için tedaviden önce ve sonra bakılan kan lipidlerinin düzeylerinin ortalaması karşılaştırılabilir. Bu çalışma döneminde her hasta, hem tedavi öncesi, hem de tedavi sonrası grupta yer alacağı için tedavi öncesi ve sonrası gruplar bağımlıdır.

ÖRNEK 41

İki farklı antihipertansif ilacın etkinliğinin birbirleriyle karşılaştırılması amacıyla bir çalışma planlanıyor. Kliniğe gelen hipertansif hastalara kayıt numaralarına bakılarak, eğer tek sayıysa A ilacı, çift sayıysa B ilacı veriliyor. Bir aylık tedavi süresi sonunda hastaların kan basıncıları ölçülüyor. İki grubun tedavi sonundaki kan basınclarının ortalamaları karşılaştırıldığında A ilacının verdiği grupta tedavi sonrası ortalama kan basıncının B ilacı verilen gruba göre daha düşük olduğu saptanıyor.

Örnek 41'deki çalışma bağımsız eşzamanlı kontrollü klinik çalışma düzennedir. Bu tür çalışmaların sonuçlarının genellikle oldukça güvenilir olduğunu biliyoruz. Ama, hastalar, tamamen tesadüfen iki gruptan birine alınırken, B grubunda daha ağır hipertansifler, A grubunda daha hafif hipertansifler toplanmış olabilir.

Bu durumda ağır hipertansiflerin çoğunlukta olduğu gruptaki B ilacı daha az etkinmiş gibi sonuç çıkabilir. Ya da bir grupta kadınlar erkeklerden daha fazla olabilir ve belki de ilaçlardan biri ya da her ikisi birden farklı cinslerde farklı şekilde etki ediyor olabilir. Ya da bir grupta şişmanlar, diyabetikler, böbrek fonksiyonu bozuk olanlar vb. gibi özellikle hastalar daha fazla olabilir ve ilaçların etkinliğinin objektif olarak karşılaştırılmasını etkileyebilirler. Bu durumda ne yapmalı?

Uygulanabilecek yöntemlerden biri, ilacın etkinliğini belirleyebilecek etkenleri önceden belirleyip, iki gruba bu etkenler açısından benzer (ya da eş) denekler almak olabilir.

Örnek 43'teki yöntemle kişiden kişiye değiştirebilecek etkenleri kontrol etmiyerek, ama belki de o ana dek bilinmeyen bir etken ilaçların etkinliğini artırmayı, ya da azaltmayı olabilir. Örneğin, kişinin egzersiz yapması ile ilaçların etkinliği artıyorsa ve gruptardan biri, diğer gruba göre daha fazla egzersiz yapan hastalardan olmuşsa, yine yaniltıcı sonuçlar alabiliriz. İki gruba alınarak hastaları kontrol ederken, göz önüne alınacak etkenlerin neler olduğu, bâzın çalışma bittikten sonra, hatta daha sonra başka çalışmaların sonuçları elde edildikten sonra anlaşılabilir.

ÖRNEK 42

Antihipertansif ilaçların etkinliğini hipertansiyonun şiddetti, hastanın yaşı, cinsiyeti, vücut ağırlığı, birlikte bulunan diğer hastalıklar ve hastanın kullandığı diğer ilaçların etkilediği bilinmektedir. Bu değişkenlerden sayısal olanları kategorilere ayırmakla işe başlayıp, hipertansiyonun şiddetini "hafif, orta, ağır", yaşı " <40 , $41\text{-}60$ ve >60 ", vücut ağırlığını da "zayıf, normal, şişman" olarak kategorize edelim. İlk gelen hasta 55 yaşında, hafif hipertansiyonu olan, şişman, ilaç kullanmayan ve başka bir hastalığı olmayan bir bayan olsun. Bu hastaya tamamen tesadüfi olarak, örneğin yazı-tura atarak seçtiğimiz A ilacını verelim. Daha sonra gelen her hastada yaş, hipertansiyonun şiddetti, vb. gibi etkenlere bakarak, daha önceki hastalarla karşılaşırıyalım ve tüm etkenler açısından eş değilse, yine tesadüfi olarak A ya da B grubuna alalım. Diyelim ki beşinci hasta 49 yaşında, hafif hipertansiyonu olan, şişman, ilaç kullanmayan ve başka bir hastalığı olmayan bir bayan, yani ilk hastaya eş. Bu hastaya hangi ilacı vereceğimizi artık yazı-turayla belirlemeyeceğiz; daha önceki hastaya, yani bu hastanın eşine A ilacı verildiğine göre, B ilacını vereceğiz. Bu şekilde çalışma tamamlandığında, iki grup hasta ilaçların etkinliğini değiştirebilecek etkenler açısından benzer hastalardan (yani eşlerden) oluşmuş olacaktır.

Kişiden kişiye değişimlecek etkenleri kontrol etmenin en iyi ve güvenilir yolu, her iki gruba da aynı denekleri almaktır. Yani çalışmaya alınan her hastaya hem A tedavisini, hem de B tedavisini uygulamaktır. Bu durumda her iki ilaç grubunda da aynı kişiler yer alacağı için ilaçların etkinliği ile ilgili en güvenilir sonuçlar elde edilmiş olur. Bu çalışma düzeneğine çapraz düzen (*cross-over design*) adı verildiğini biliyorsunuz.

Bazen de bir tedavinin ya da etkenin etki etmesi beklenen değişkendeki değişimlerin miktarını saptamak için çalışma planlanabilir.

ÖRNEK 43

Hemodiyalize alınan hastaların hemodiyaliz öncesi ve sonrasında beyaz küre değerlerine bakarak, hemodiyalizle beyaz küre değerlerinin değişip değişmediği araştırılıyor. Bu çalışmada hastaların hemodiyaliz öncesi ve sonrası beyaz küre değerlerinin oluşturdukları gruplar, bağımlı gruplardır.

Örnek 42 ve 43'teki çalışmalar bağımlı gruplar (ya da örneklemeler) içeren çalışmalarlardır. Örnek 42'deki eşleştirilmiş düzende gruplardaki hastalar farklı kişiler olduğu halde birbirlerine benzer olduğu için gruplar bağımlı olarak varsayıılır.

Bağımlı grup ortalamaları bağımlı örneklem t testi ya da diğer adı ile Student bağımlı örneklem t testi ile karşılaştırılır. Bu testte iki grubun ortalamaları arasındaki farkın, deneklerin değerlerinin farklarının ortalamasına eşit ol-

ması gerçeği kullanılır. Yani aslında karşılaştırılan iki grubun ortalamalarının birbirlerinden farklı olup olmadığı değil, iki grup değerleri arasındaki farkların ortalamasının "0"dan farklı olup olmadığıdır.

ÖRNEK 44

Kronik böbrek yetmezliğine bağlı aneminin tedavisi amacıyla eritropoietin tedavisi verilen 12 hastada tedaviden önce ve 3 ay sonra hemoglobin (Hb) düzeyleri bakılıyor. Hastalardaki Hb değerleri Tablo 1'de verilmiştir.

Tablo 1 Örnek 44'teki çalışmanın verileri.

Tedavi öncesi Hb (g/dL)	Tedavi sonrası Hb (g/dL)	Hb'deki değişme miktari (g/dL)
7.5	12.3	4.8
7.6	10.9	3.3
8.6	9.8	1.2
9.3	13.4	4.1
7.8	11.1	3.3
9.8	9.3	-0.5
9.3	9.9	0.6
10.3	12.3	2.0
8.4	11.3	2.9
7.0	9.0	2.0
11.9	10.9	-1.0
7.5	12.3	4.9
X : 8.7	X : 11.0	X : 2.3
SD: 1.4	SD: 1.4	SD: 1.9

Tedavi öncesi Hb ortalaması 8.7 g / dL iken, tedavi sonrasında 11.0 g / dL'ye yükselmiş, yani ortalama 2.3 g / dL kadar artmıştır. Dikkat edilirse, farkların ortalaması, ortalamaların farklarına eşitken, standart sapmalar için bu durum söz konusu değildir.

Sonuç olarak her hastanın Hb değişme miktarının ortalaması ve standart sapması bilindiğine göre, bu grubun Hb değişme miktarının, yani Tablo 1'de en sağdaki sütunda bulunan değerlerin 0'dan farklı olması durumunda onç ve sonra gruptarının ortalamalarının birbirinden farklı olduğu söylenebilir.

Güven sınırları

Tek grup ortalamasının bilinen bir değerden farklı olup olmadığını testinde olduğu gibi bağımlı iki grup ortalaması arasındaki farkın güven sınırlarının hesaplanabilmesi için de verilerin normal dağılım göstermesi koşulu vardır. Yalnız burada gruptardaki verilerin dağılımı değil, her denekte önce sonra arasındaki farkların oluşturduğu verilerin dağılımı söz konusudur. Çünkü hesaplamların yapılacakları veriler, grupların ortalamalarını oluşturan ham veriler değil, denek değerleri arasındaki farklardır.

Tek grup ortalamasının güven sınırlarının belirlenmesinde değer ortalamaları ve ortalamanın standart hatasının (SEM) kullanıldığı yukarıda anlatılmıştır. İki grup ortalaması arasındaki farkın güven sınırlarında da aynı formül kullanılır. Yalnız değerlerin ortalaması ve standart hata yerine farkların ortalaması ve standart hatası (SE_d) kullanılır.

$$\%95 \text{ güven sınırları} = \text{Farkların ortalaması} \pm (t_{n-1, \alpha}) (SE_d)$$

$$SE_d = \frac{SD_d}{\sqrt{n}}$$

n = Denek sayısı

Tablo 1'deki verileri formüldeki yerine koyunca eritropoietin öncesi ve sonrası Hb değerleri arasındaki farkın güven sınırları $2.3 \pm (2.201)(1.9 / \sqrt{12})$, yani alt güven sınırı 1.1, üst güven sınırı 3.5 g/dL'dir. Yani %95 olasılıkla, ya da başka bir deyişle %5 yanılma payıyla Hb değerindeki değişme miktarı 1.1 ile 3.5 g/dL arasındadır. Alt sınır 0'dan büyük olduğu, ya da güven aralığı 0'ı içermediği için, farkların ortalamasının 0'dan farklı olduğu, yani eritropoietin öncesi ve sonrası Hb değerlerinin ortalamalarının aynı olmadığı sonucuna varılabilir.

Hipotez testi (bağımlı iki grup ortalaması için Student t testi)

Bağımlı iki grup ortalamasının hipotez testi ile karşılaştırılmasında da aslında yapılan, deneklerin önce ve sonra değerleri arasındaki farkların ortalamalarının 0'dan farklı olup olmadığını test edilmesidir.

Hipotez testi sırasında t değeri aşağıdaki formülle hesaplanır.

$$t_{n-1, 0.05} = \frac{\text{Farkların ortalaması}}{\text{SE}_d}$$

$$\text{SE}_d = \frac{\text{SD}_d}{\sqrt{n}}$$

n = Denek sayısı

Yine Örnek 44'teki verileri kullanırsak, hesaplanan t değeri $2.3 / (1.9 / \sqrt{12}) = 4.19$ olacaktır. Bu değer, serbestlik derecesi 11 ve yanılma düzeyi 0.05'e karşılık gelen tablo değeri olan 2.201'den büyük olduğu için, %5 yanılma olasılığı ile önce ve sonra Hb değerleri arasındaki farkın 0'dan farklı olduğu anlaşılır.

Verilerin dağılımı normal değilse

Bağımlı iki grup ortalamasının karşılaştırılmasında t testinin kullanılabilmesi için denek değerleri arasındaki farkların dağılıminin normal dağılıma yakın olması gereklidir. Denek sayısı 30'dan fazla ise denek değerlerinin normal dağıldığı varsayılabılır. Daha az denek varsa ya da dağılımin normal olmadığını kuşkulanılıyorsa, grafik yöntemlerle ya da gerekli testlerle dağılım özellikleri araştırılmalıdır (Bölüm 3'e bakınız).

Eğer deneklerin dağılımı normal değilse, tek grup ortalamasının test edilmesinde olduğu gibi değişim uygulanabilir. Alternatif olarak normal dağılım göstermeyen iki bağımlı grup ortalamasının karşılaştırılmasında t testinin nonparametrik olmayan karşılıkları kullanılabilir. Bu testler içinde en yaygın olarak kullanılan Wilcoxon işaret testidir.

Wilcoxon testinin ana ilkesi, değerlerin kendilerinin değil, sıra numaralarının kullanılmasıdır, yani aslında değerlere sıra dönüşümü uygulanır.

Wilcoxon testinde, bağımlı iki grubun ortalamaları değil, ortancaları arasındaki farkın önemli olup olmadığı test edilir.

Önce ve sonra değerleri arasındaki farkların mutlak değerleri küçükten büyüğe sıraya konup, sıra numarası verilir. Tablo 2'de ilk sütunda Hb'deki değişme miktarı, ikinci sütunda ise Hb'deki değişme miktarının mutlak değerleri görülmektedir. Bu sayılar içinde en küçüğü 0.5 olduğu için sıra numarası "1" olmalıdır. Bunu izleyen 0.6'ya "2", 1.0'a "3" ve 1.2'ye "4" sıra numaraları verilmiştir. Beşinci ve altıncı en küçük değerler olarak iki tane 2.0 ol-

duğu için bu iki değere sıra numarası olarak 5 ve 6'nın ortalaması olan "5.5" verilmiştir. Tüm değerlere sıra numaraları verildikten sonra, değişme miktarı 0'dan küçük olanlar, yani işaret "- " olanlar ve 0'dan büyük olanlar, yani işaret "+ " olanlar ayrı ayrı toplanır. Tablo 2'deki verilere dayanarak "- " işaretli değerlerin toplamının 4, "+" işaretli değerlerin toplamının 74 olduğu görülmektedir. Daha sonra bu sıra numaraları toplamları uygun formüller ve tablolardan yardım ile karşılaştırılır.

Örnek 44'te denek sayısı 30'dan az olduğu için, t testi yerine Wilcoxon testi uygulanması daha doğrudur. T testinde hesaplanan p değeri 0.002 iken, Wilcoxon testi ile hesaplanan p değeri 0.006'dır, bu nedenle yorum her iki testte de farklı değildir.

Genel olarak, normal dağılım göstermeyen değerler için Wilcoxon testi, t testine göre daha güçlündür, yani önce ve sonra değerleri arasında fark varsa, daha doğru olarak saptar. Örneğin normal dağılım göstermeyen değerler için t testi kullanıldığında p değeri 0.08 gibi sınırsız bulunurken, Wilcoxon testi uygandığında, 0.05'ten küçük p değeri bulunabilir.

Tablo 2 Örnek 44'teki çalışmanın verilerine sıra dönüşümü uygulanması sırasındaki adımlar.

Hb'deki değişme miktarı (g/dL)	Hb'deki değişme miktarının mutlak değeri	Sıra no
4.8	4.8	11
3.3	3.3	8.5
1.2	1.2	4
4.1	4.1	10
3.3	3.3	8.5
-0.5	0.5	- 1
0.6	0.6	2
2.0	2.0	5.5
2.9	2.9	7
2.0	2.0	5.5
-1.0	1.0	3
4.9	4.9	12

Normal dağılım gösteren değerler için her iki testin gücü aynıdır. Bu nedenle Wilcoxon testi tip alanında giderek artan sıklıkta kullanılmaktadır.

Wilcoxon testinde hepsi aşağı yukarı aynı sonucu veren çeşitli hesaplama yöntemleri vardır. Bu nedenle farklı bilgisayar programlarında çok az da olsa birbirinden farklı p değerleri elde edilebilir.

BAĞIMSIZ İKİ GRUP ORTALAMASININ KARŞILAŞTIRILMASI

Eğer iki gruptaki değerler farklı deneklerden elde edilmişse, yani deneklerin hiçbirini iki grupta birden yer alımırsa, gruplar birbirinden bağımsızdır.

Örneğin eritropoietinin anemi tedavisindeki etkisini araştırmak için, bir grup hastada tedavi öncesi ve sonrası Hb değerlerini karşılaştırmak yerine, eritropoietin alan ve almayan iki grubun Hb değerlerinin karşılaştırılacağı bir çalışma düzeni seçilebilir. Bu durumda bağımsız iki grup var demektir.

Bağımsız iki grup ortalamasının karşılaştırılmasında da t testi kullanılır. Bu durumda da her t testinde olduğu gibi değerlerin dağılımının normal olması varsayılmıştır. Yine her iki gruptaki denek sayıları 30'dan fazla ise bu koşulun sağlandığı varsayılabılır. Daha önceki t testlerinden farklı olarak, bir diğer varsayılmış ise iki gruptaki standart sapmaların ya da varyanslarının da eşit (ya da homojen) olması gerekliliğidir.

İki gruptaki denek sayıları aşağı yukarı eşitse, varyansların eşitliği koşulu aranmayıabilir. Ama grplardaki denek sayıları birbirinden farklı ise, bu koşulun sağlanması gerekmektedir. Bu nedenle çalışma düzenini kurarken, grplardaki denek sayılarının birbirine eşit, ya da en azından yakın olması planlanmalıdır.

Güven sınırları

Bağımsız iki grubun ortalamaları arasındaki farkın güven sınırlarının formülü, daha önceki formüllere göre daha karmaşık olduğu için burada verilmemiştir. Ancak ortalamalar arasında fark olup olmadığını karşılaştırırken, yorum benzerdir.

Ortalamalar arasındaki farkın %95 güven aralığı 0'ı içeriyorsa, ortalamalar birbirinden farklı değildir; 0'ı içermiyorsa, farklıdır.

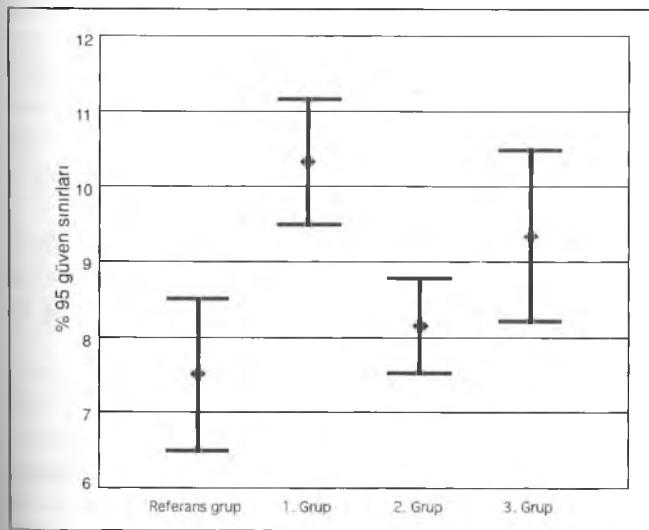
ÖRNEK 45

Örnek 44'teki çalışmada bir grup hastada eritropoietin öncesi ve sonrası Hb ölçümü yapıp karşılaştırmak yerine, birine eritropoietin, diğerine placebo verilen bağımsız iki grup kullanılıyor. Tablo 1'de birinci sütundaki değerlerin placebo verilen, ikinci sütundaki değerlerin ise eritropoietin verilen gruba ait olduğunu varsayıyalım. Bu değerlere göre hesaplanan iki grup ortalaması arasındaki farkın %95 güven sınırları 1.1 ve 3.5 g/dL'dır. Güven aralığı 0'ı içermediği için grup ortalamaları birbirinden farklıdır. Burada güven sınırları tespitlenen bağımlı ve bağımsız örneklem için aynı çıkmıştır. Genellikle bağımsız grpların ortalamaları arasındaki farkın güven aralığı, bağımlı grplara göre daha genişlit

Gruplardaki denek sayıları 10 ya da daha fazlaysa, grupların ayrı ayrı %95 güven sınırları hesaplandığında da ortalamalar arasındaki fark hakkında bir yorum yapılabilir.

Bağımsız iki grubun ortalamalarının güven sınırları şu üç durumdan birine uyar:

1. Bir grubun %95 alt güven sınırı, diğer grubun %95 üst güven sınırından daha yüksek olabilir. Bir diğer deyişle, iki grubun güven aralıkları kesişmeyebilir. Bu durumda iki grup ortalamalarının birbirinden farklı olduğu sonucuna varılabilir. Şekil 1'de referans grubun Hb değerinin %95 güven sınırları 6.5 ve 8.5, birinci grubun %95 güven sınırları 9.5 ve 11.2'dir. Bu iki grubun ortalamaları %5'den daha az yanılma olasılığı ile birbirinden farklıdır.
2. Bir grubun ortalaması, diğer grubun %95 güven aralığı içinde olabilir. Bu durumda iki grup ortalamalarının birbirinden farklı olmadığı sonucuna varılabilir. Şekil 1'de referans grubun Hb değerinin %95 güven sınırları 6.5 ve 8.5, ikinci grubun Hb ortalaması ise 8.2'dir. Bu iki grubun ortalamaları %5'den daha az yanılma olasılığı ile farksızdır.
3. Her iki grubun ortalaması da diğer grubun %95 güven aralığının dışında olup, güven aralıkları kesişebilir. Bu durumda yorum yapılamaz. Hipotez testi yapılmalıdır. Şekil 1'de referans grubun Hb değeri ortalaması 7.5 ve %95 güven sınırları 6.5 ve 8.5, diğer grubun Hb değeri ortalaması 9.4 ve %95 güven sınırları 8.2 ve 10.5'tür. Bu iki grubun Hb ortalamaların farklı olup olmadığı hipotez testi ile test edilmelidir.



Şekil 1
Bağımsız iki grup ortalamasının güven aralıklarının karşılaştırılması. Grup ortalamaları ◆ işaretleri ile, güven sınırları ise dikey çizgilerin alt ve üst sınırlarındaki kısa yataç çizgilerle gösterilmiştir. Referans grupta karşılaştırıldığında 1'inci grubun ortalaması farklıdır, 2'inci grubun ortalaması farklı değildir, 3'üncü grup için ise kesin bir yorum yapılamaz.

Makalelerde kullanılan grafiklerde sıklıkla ortalama ile birlikte standart hata gösterilir. İki grubun ortalaması böyle bir grafikte yan yana görüldüğünde, dikkat edilmezse yanıltıcı olabilir. Çünkü ortalama \pm standart hata, çalışma grubunun seçildiği evrenden seçilecek aynı büyüklükteki olası örneklemelerin ortalamalarının yaklaşık %68'ini içerir. Yani aşağı yukarı %68 güven aralığına eşittir. Oysa, iki grubun ortalamaları arasında fark olduğu sonucuna varmak için %95 güven aralığı kullanılmalıdır.

Bu nedenle bu amaçla verilen grafiklerde, aşağı yukarı %95 güven aralığına eşit olan 2SEM, yani standart hatanın iki katı kullanılmalıdır.

Hipotez testi (bağımsız iki grup ortalaması için Student t testi)

Bağımsız iki grubun ortalamaları arasında fark olup olmadığını diğer bir Student t testi ile anlayabiliriz. Yine gruptardaki değerlerin normal dağılımı varsayımları her parametrik teste olduğu gibi bu durumda da geçerlidir ve yine her gruptaki denek sayısı 30 ya da daha fazla ise normal dağılımı varsayılabılır.

Bu teste de t değeri hesaplanır ve tablodaki t değeri ile karşılaştırılır. Formülü diğer testlere göre daha karışık olduğu için verilmemiştir. Bir gruptaki denek sayısı n_1 , diğer gruptaki denek sayısı n_2 ise, tabloda serbestlik derecesi olarak " $n_1 + n_2 - 2$ ", yarılma düzeyi olarak 0.05'e karşılık gelen t değerine bakılmalıdır. Örneğin yukarıdaki örnekte iki grupta 12'serden toplam denek sayısı 24 olduğu için tabloda 22 serbestlik derecesine karşılık gelen t değeri ile hesaplanan t değeri karşılaştırılmalıdır. Hesaplanan t değeri, tablo değerinden büyükse, iki grubun ortalamalarının birbirinden farklı olduğu sonucuna varılabilir. Eritropoietin örneğinde hesaplanan t değeri 4.01 ve p değeri 0.001 olduğu için eritropoietin tedavisi verilen grubun Hb ortalamasının, verilmeyen grubun Hb ortalamasından daha yüksek olduğu söylenebilir.

Grup varyanslarının eşitliği

Bağımsız iki grubun ortalamalarının karşılaştırılmasında t testi uygulanıp bilmesi için gruptarın varyanslarının eşit olması gereği belirtilmiştir. Bu koşul, özellikle gruptardaki denek sayıları eşit değilse, daha da önemlidir. Bu nedenle t testine geçmeden önce varyansların eşitliği test edilmelidir.

Varyansların eşitliği (homojenliği) için en sık kullanılan test, Levene testi ya da kısaca F testi olarak bilinir. Bu teste büyük olan varyansın küçük olan varyansa bölümü F değerini verir. Varyansı büyük olan gruptaki denek sayısı n_1 , diğer gruptaki denek sayısı ise n_2 ise, hesaplanan F değeri, F tablosunda $n_1 - 1$ ve $n_2 - 1$ 'e karşılık gelen F değeri ile karşılaştırılır. F tablosunda $n_1 - 1$ e

karşılık gelen sütünla, $n_2 - 1$ 'e karşılık gelen sıranın kesiştiği yerdeki değer, tablodaki F değeridir. Eğer hesaplanan F değeri, tablodaki F değerinden büyüğse, $p < 0.05$ 'dir ve varyanslar eşit değildir.

Eğer iki grubun varyansları eşit değilse, t değerinin hesaplanması ile ilgili bazı değişiklikler yapılır. Bu durumda iki grubun ortalamalarının farklı bulunabilmesi, varyansların eşit olması durumuna göre zorlaşır. Örneğin varyansları eşit iki grubun ortalamaları 7.5 ve 9.5 iken ortalamalar birbirinden farklı bulunurken, aynı ortalamaları olan, ama varyansları eşit olmayan iki grubun ortalamaları farklı bulunmayabilir.

Student t testi varsayımları sağlanamıyorsa

Dağılımı normal olmayan bağımsız iki grup varsa, değerlere dönüşüm uygulandığında, dönüştürülmüş değerler normal dağılıyorsa ve yine bu dönüştürülmüş değerlerin varyansı eşitse, t testi uygulanabilir. Varyansların eşit olmadığı, ama dağılımların normal olduğu durumlarda ise, düzeltilmiş t testi uygulanabilir.

Bu iki varsayımdan herhangi birinin sağlanamadığı durumlarda uygulanailecek bir diğer yöntem ise bağımsız iki grup ortalaması için olan t testinin nonparametrik karşılığı olan Mann-Whitney U testi'dir. Bağımlı iki grup ortancalarının karşılaştırıldığı nonparametrik test olan Wilcoxon testinde olduğu gibi, Mann-Whitney U testinde de aslında yapılan değerlere sıra dönüşümü uygulanması ve ortalamalar yerine ortancaların karşılaştırılmıştır. İki grubun değerleri karışık olarak küçükten büyüğe sıralanır ve sıra numarası verilir. Daha sonra bu sıra numaraları kullanılarak t değeri hesaplanır. Hesaplanan t değeri $n_1 + n_2 - 2$ serbestlik derecesi ve 0.05 yanılma düzeyindeki t değerinden büyükse, iki grubun ortancalarının birbirinden farklı olduğu anlaşılır.

Normal dağılıma uygun olmayan verilere t testi uygulanırsa,其实e fark olduğu halde, test sonucunda fark bulunamayabilir. T testi varsayımlarının sağlanamadığı durumlarda Mann-Whitney U testi, t testinden daha güçlündür.

SPSS'TE BAĞIMLI İKİ GRUP ORTALAMASI NASIL KARŞILAŞTIRILIR?

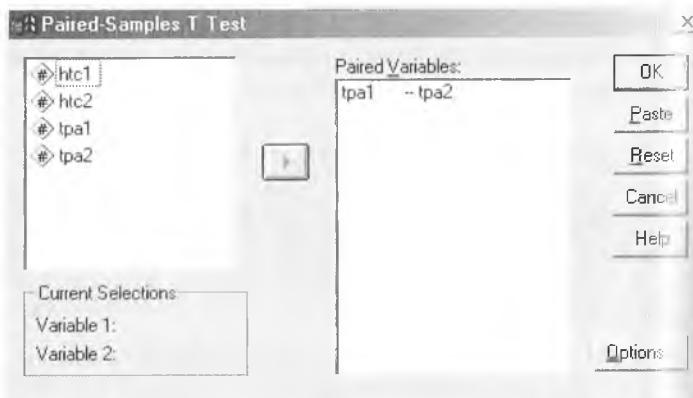
Student t testi

Analyze >> Compare Means >> Paired-Samples T Test

Yukarıdaki seçenekle Paired-Samples T Test diyalog kutusu açılır (Şekil 2). Bildikleri listede dosyadaki sayısal değişkenler yer almaktadır. Değişken çiftlemeni seçmek için önce değişkenlerden birini kliklediğinizde, bu değişken adı

Current Selections'daki Variable 1'e aktarılacaktır. İkinci değişkeni kliklediğinizde Variable 2'ye bu değişken aktarılacaktır.

Daha sonra, bu değişken çifti sağdaki *Paired Variables* listesine aktarılır. *Paired Variables* listesine birden çok değişken çifti seçebilirsiniz.



Sekil 2
SPSS'te "Paired-Samples T Test" diyalog kutusu.

ÖRNEK 46

Venöz oklüzyon testi (VOT), kola tansiyon aletinin manşonunu bağlayıp, sistolik ve diystolik basınç değerleri arasında 10-15 dakika sabit tutularak yapılan bir test

tir. Oklüzyon yapılan koldaki venlerden oklüzyondan önce ve sonra alınan bazı değerler değişiklik gösterir. Örneğin Hb ve Htc değerleri bir miktar artar, ayrıca endotelden salınan bazı maddelerin salınımını ve dolayısıyla plazma düzeyleri artar. Endotelde hastalık varsa, VOT'ne yanıt bozulur. Bir çalışmada 54 sağlam kişiye VOT yapılarak, normal kişilerde endotelden salınan doku plazminojen aktivatörü (tPA) plazma düzeylerinde değişiklik olup olmadığıının belirlenmesi amaçlanıyor

VOT öncesi ölçümler tpa1, VOT sonrası ölçümler tpa2 adlı değişkenler kaydedilmiş. VOT ile tPA miktarında değişme olup olmadığı "iki bağlılı grup t testi" ile yapılmıştır. Sonuçlar Şekil 3'de verildiği gibidir.

Şekil 3'de *Paired Samples Test* tablosu "t testi" sonuçlarını vermektedir. Mean olarak iki ortalamanın farkı, ya da önce-sona ölçümleri arasındaki farklıların ortalaması verilmektedir. (İkisi de aynı rakamı verir). Sig 2-tailed'm altı daki değer, p değeri <0.05 olduğu için (0.000 olarak verilen)

değerinin en azından <0.001 olduğu anlaşılmaktadır), tpa1 ile tpa2 ortalamaları arasındaki fark istatistiksel olarak önemlidir.

İlk değişken olarak tpa2, ikinci değişken olarak tpa1 seçilmiş olsaydı, farklıların ortalaması -6.930 değil, $+6.930$, ve yine t değeri de -11.705 değil, $+11.705$ olacaktı, ama hesaplanan p değeri değişimeyecekti. Değişken çiftlerinin hangi sırayla seçildiğinin önemi yoktur.

Paired Samples Statistics ve *Paired Samples Correlation* tablolarında önce ve sonra değerlerinin ortalama, standart sapma ve standart hata değerleri ve iki değişken arasındaki bağıntının derecesi ve önemi de verilmektedir.

Paired Samples Statistics				
	Mean	N	Std. Deviation	Std. Error Mean
Pair 1	TPA1	11.79	54	3.66
	TPA2	18.72	54	.50
			7.47	1.02

Paired Samples Correlations				
	N	Correlation	Sig.	
Pair 1	TPA1 & TPA2	54	.920	.000

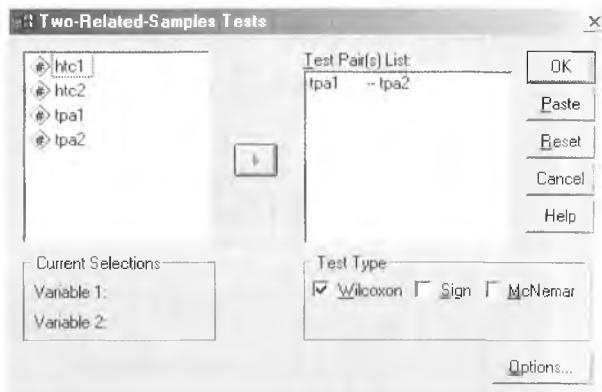
Paired Samples Test											
	Paired Differences					t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference							
				Lower	Upper						
Pair 1	TPA1 - TPA2	-6.93	4.35	.59	-8.12	-5.74	-11.705	.000			

Şekil 3 VOT öncesi ve sonrası tPA ortalamalarının t testi ile karşılaştırılması.

Wilcoxon işaret testi

Analyze >> Nonparametric Tests >> 2 Related Samples

Yukarıdaki seçimle Two-Related Samples Tests diyalog kutusu açılır (Şekil 4). Soldaki listede dosyadaki sayısal değişkenler yer almaktadır. Değişken çiftlerini seçmek için önce değişkenlerden birini kliklediğinizde, bu değişken adı Current Selections'daki Variable 1'e aktarılacaktır. İkinci değişkeni kliklediğinizde Variable 2'ye bu değişken aktarılacaktır. Daha sonra, bu değişken çifti sağdaki Test pairs listesine aktarılır. Paired variables listesine birden çok değişken çifti seçebilirsiniz.



Şekil 4
SPSS'te "Nonparametric tests" seçenekinden "Two-Related Samples Tests" diyalog kutusu.

Bu diyalog kutusu ilk açıldığında Wilcoxon testi seçilmiş durumdadır. Wilcoxon testine alternatif olarak işaret testini *Sign* seçenekinden yapabilirsiniz. Ancak Wilcoxon testine göre hiç bir üstünlüğü olmadığı gibi, daha gücsüz olması nedeniyle işaret testinin kullanılması yararsızdır. Diğer seçenek McNemar testi ise, bağımlı iki grup oranlarının karşılaştırıldığı testlerdir. Bu testlerle ilgili ayrıntılar Bölüm 12'de anlatılmıştır.

Örnek 42'deki çalışmada eğer parametrik test varsayımlarından dağılımin normal olması koşulu sağlanamıyorsa, Wilcoxon testi ile grup ortalamaları karşılaştırılmalıdır. Şekil 5'de tpa1 ve tpa2 ortalamalarının Wilcoxon testi ile karşılaştırılmasının sonuçları görülmektedir. Ellidört hastanın üçünde tPA2'nin tPA1'den daha düşük, 51'inde tPA2'nin tPA1'den yüksek olduğu görülmektedir. TPA2 değeri tPA1'e eşit olan hasta yoktur. Sig. 2-tailed'deki p değeri <0.05 (0.000) olduğuna göre tPA1 ile tPA2 ortancalarının eşit olmadığı sonucuna varılmıştır.

Ranks			
	N	Mean Rank	Sum of Ranks
TPA2 - TPA1 Negative Ranks	3 ^a	10.17	30.50
Positive Ranks	51 ^b	28.52	1454.50
Ties	0 ^c		
Total	54		

a. TPA2 < TPA1
 b. TPA2 > TPA1
 c. TPA1 = TPA2

Test Statistics^b

	TPA2 - TPA1
Z	-6.138 ^a
Asymp. Sig. (2-tailed)	.000

a. Based on negative ranks.
 b. Wilcoxon Signed Ranks Test

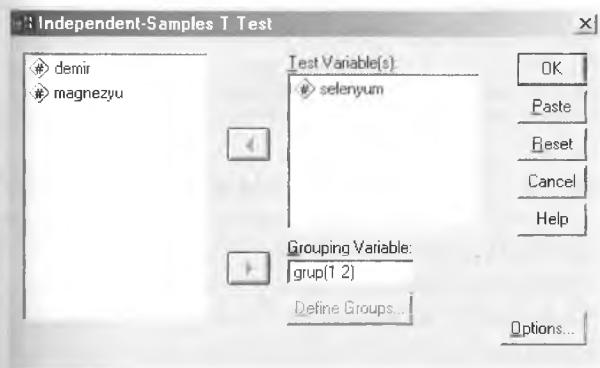
Şekil 5
VOT öncesi ve sonrası tPA ortancalarının Wilcoxon testi ile karşılaştırılması.

SPSS'TE BAĞIMSIZ İKİ GRUP ORTALAMASI NASIL KARŞILAŞTIRILIR?

Student t testi

Analyze >> Compare Means >> Independent-Samples T Test

Yukarıdaki seçimle **Independent-Samples T Test** diyalog kutusu açılır (Şekil 6). Burada bağımsız değişken *Grouping variable* kutusuna, ortalamaları karşılaştırılacak değerleri temsil eden değişken, yani bağımsız değişken ise *Test variable(s)* kutusuna seçilmelidir. *Grouping variable* kutusuna alınan değişkenin hangi değerlerinin grupları ayırdığı ya da belirlendiği *Define groups* seçenekü ile belirlenmelidir.



Şekil 6
SPSS'te "Independent-Samples T Test" diyalog kutusu.

Akut miyokard infarktüsü (AMI) geçirmiş hastalarda ve normal kontrollerde serumda selenyum (Se) düzeyi ölçülüyor. İki gruptaki Se ortalamaları karşılaştırılıyor.

Örnek 47'deki işlemi yapmak için, hastaların grubunu girdiğimiz değişkeni (grup) *grouping variable* olarak seçmeliyiz. AMI grubu "1", kontrol grubu "2" ile kodlanmışsa, *Define groups* seçenekinde grupları 1 ve 2 olarak tanımlamalıyız. *Options* penceresinden önceden %95 olarak seçilmiş olan güven aralığını değiştirmeye olanağı vardır.

Sonuçta Şekil 7'deki tablo elde edilir. *Group Statistics* tablosunda her iki grubun selenyum değerlerinin ortalama, standart sapma (SD) ve standart hata (*SE of mean*) değerleri verilmiştir. *Independent Samples Test* tablosunda ise varyansların homojen olup olmadığını test edildiği Levene testinin sonuçları vardır.

Şekil 7
Örnek
47'deki çal-
ışmada
akut miyo-
kard infark-
tüsü ve nor-
mal grupta
serum se-
lenyum or-
talamaları-
nın karşıla-
şırıldığı
t testi
sonuçları.

Group Statistics					
	GRUP	N	Mean	Std. Deviation	Std. Error Mean
SELENYUM	1 AMI	17	60.77	26.51	6.43
	2 Normal	18	65.06	27.43	6.47

	Independent Samples Test							
	Levene's Test for Equality of Variances		t-test for Equality of Means					
	F	Sig.	t	df	Sig. (2-tailed)	95% Confidence Interval of the Difference		
SELENYUM	Equal variances assumed	.015	.902	-.470	33	.641	-2.11	-1.39
	Equal variances not assumed			-.471	33.0	.641	-2.11	-1.39

Levene testi ile elde edilen p değeri >0.05 ($=0.90$) olduğu için varyansların homojen (eşit) olduğu sonucuna varılabilir. T testi varyansların eşit olduğu ve olmadığı her iki durum için de uygulanmıştır. Bu örnekte varyanslar eşit olduğu için *Equal variances assumed* satırındaki değerleri alabiliriz.

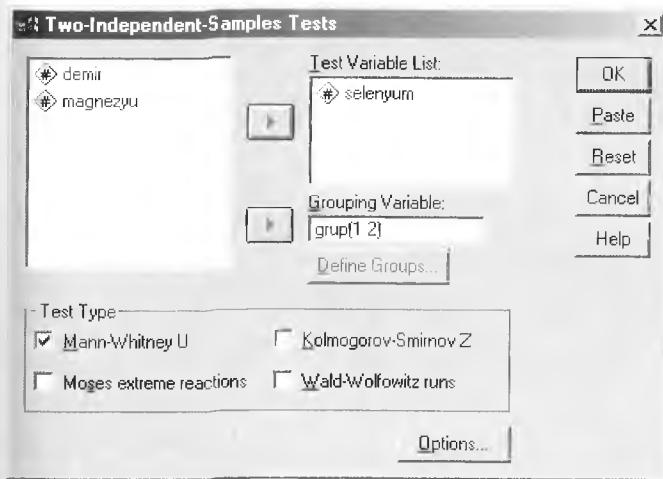
Sig 2-tailed sütunundaki değerler, p değerleridir. P değeri 0.64, yani >0.05 , olduğuna göre Se ortalamalarının iki grup arasında farklılık gösterdiğini söyleyebiliriz.

leme olası değildir. En sağdaki 95% *Confidence Interval of the Difference* sütunu ise iki grup arasındaki farkın %95 olasılıkla -22.86 ve +14.28 arasında olduğu görülmektedir. Güven aralığı 0'ı (sıfır) içerdiği için %95 olasılıkla iki grup Se ortalamalarının farklı olmadığı anlaşılmaktadır.

Mann-Whitney U testi

Analyze >> Nonparametric Tests >> 2 Independent Samples

Yukarıdaki seçimle Two-Independent-Samples Tests diyalog kutusu açılır (Şekil 8). Burada bağımsız değişken *Grouping variable* kutusuna, ortalamaları karşılaştırılacak değerleri temsil eden değişken, yani bağımsız değişkenler ise *Test variable* list kutusuna seçilmelidir. *Grouping variable* kutusuna alınan değişkenin hangi değerlerinin grupları ayırdığı ya da belirlediği *Define groups* seçenekü ile belirlenmelidir.



Şekil 8
SPSS'te "Nonparametric tests" seçenekinden "Two-Independent-Samples Tests" diyalog kutusu.

Bu diyalog kutusu ilk açıldığında Mann-Whitney U testi seçilmiş durumdadır. Mann-Whitney U testine alternatif olarak nadiren kullanılabilen Kolmogorov-Smirnov testi, Moses testi ve Wald-Wolfowitz testi de seçilebilir. Bu testlerin Mann-Whitney U testine üstünlükleri yoktur ya da başka amaçlar için kullanılabilir.

Yukarıdaki örnekte grplardaki dağılım normal değilse, Mann-Whitney U testi uygulanmalıdır. Analiz sonucu Şekil 9'da gösterilmiştir. *Sig 2-tailed*'de gösterilen p değeri >0.05 (0.67) olduğu için iki grup Se ortancalarının farklı olmadığı sonucuna varılabilir.

Şekil 9
 Örnek 47'deki
 çalışmada akut
 miyokard infarktüsü
 ve normal grupta serum
 selenyum ortancalarının
 karşılaştırıldığı
 Mann-Whitney U testi
 sonuçları.

Ranks			
GRUP	N	Mean Rank	Sum of Ranks
SELENYUM 1 AMI	17	17.24	293.00
2 Normal	18	18.72	337.00
Total	35		

Test Statistics ^b	
	SELENYUM
Mann-Whitney U	140.000
Wilcoxon W	293.000
Z	-.429
Asymp. Sig. (2-tailed)	.668
Exact Sig. [2*(1-tailed Sig.)]	684 ^a

a. Not corrected for ties.
 b. Grouping Variable: GRUP

bölüm 11

İKİDEN FAZLA ORTALAMANIN KARŞILAŞTIRILMASI

Tek yönlü varyans analizi

Varyans analizinin mantığı

Varyans analizindeki terimler

Varyans analizinin varsayımları

Varyanslar eşit olmazsa ne olur?

İki yönlü varyans analizi

Çoklu karşılaştırma yöntemleri

Önceden planlanmış (a priori) karşılaştırmalar

Sonradan yapılan (posteriori: post hoc) karşılaştırmalar

Tukey HSD yöntemi

Scheffe yöntemi

Newman-Keuls yöntemi

Dunnett yöntemi

Varyans analizinde diğer düzenler ve yöntemler

Randomize blok düzeni

Latin kare düzeni

Tekrarlı ölçümler düzeni

Nonparametrik varyans analizi

Kruskal-Wallis tek yönlü varyans analizi

Friedman iki-yönlü varyans analizi

SPSS'te varyans analizi nasıl yapılır?

Tek yönlü varyans analizi

İki yönlü varyans analizi

Tekrarlı ölçümler varyans analizi

Kruskal-Wallis tek yönlü varyans analizi

Friedman iki yönlü varyans analizi

Bağımlı ya da bağımsız iki grubun ortalamalarının karşılaştırılmasında t testlerinin kullanıldığı daha önceki bölümlerde anlatılmıştır. Öte yandan bir çok klinik araştırmada ikiden çok grup kullanılır ya da aynı gruptan ikiden çok gözlem ya da ölçüm yapılır. İlk durumda ikiden çok bağımsız grup, ikinci durumda ikiden çok bağımlı grup var demektir. Buradaki "grup" terimi, denek grubu ya da hasta grubu değil, veri grubu anlamında kullanılmaktadır.

Bu nedenle ikiden çok bağımlı gruptan söz edildiği zaman, aslında tek bir denek grubu vardır, ama bu grupta ikiden çok kez ölçüm yapılarak elde edilen veriler ikiden çok “veri grubu” oluşturmaktadır.

Madem iki grup olduğu zaman, uygun Student t testi ya da nonparametrik karşıtlıklarından Mann-Whitney U ya da Wilcoxon testleri ile grup ortalamaları karşılaştırılabilir, o halde ikiden çok grubun ortalamalarının karşılaştırılmasında grupların ikişerli olarak bu testlerle karşılaştırılmaları, ilk başta mantıklı ve geçerli gibi görünmektedir. Bu yaklaşımın其实 neden yanlış olduğunu Örnek 48'de görelim.

ÖRNEK 48

Primer akciğer kanseri, plevra metastazı, parapnömonik effüzyon ve tüberküloza bağlı pleural effüzyonu olan 20'ser hastada pleural sıvıda adenozin deaminaz (ADA) düzeyleri ölçülüyor. Bu dört bağımsız gruptaki ortalama ADA düzeyleri ikişer ikişer Student t testi ile karşılaştırılıyor ve şu sonuçlar elde ediliyor:

1. Primer akciğer kanseri ile plevra metastazı $p>0.05$

2. Primer akciğer kanseri ile parapnömonik effüzyon $p<0.05$

3. Primer akciğer kanseri ile tüberküloz $p<0.05$

4. Plevra metastazı ile parapnömonik effüzyon $p<0.05$

5. Plevra metastazı ile tüberküloz $p<0.05$

6. Parapnömonik effüzyon ile tüberküloz $p<0.05$

Altı karşılaştırmadan yalnız birincisinde iki grup arasında fark bulunmadıken, diğer beş karşılaştırmada ele alınan iki gruptan ADA ortalamalarının farklı olduğu sonucuna varılmıştır.

Her bir karşılaştırma için yanılma payı %5'dir, yani “iki grup ADA ortalaması birbirinden farklıdır” derken, %5 olasılıkla hata yapılması göze alınmalıdır. Bu durum, beş karşılaştırma için de geçerli olduğu için, dört grubun ADA ortalamaları hakkında global olarak aşağı yukarı %25 olasılıkla yanılma şansının hesaplanması burada olduğu gibi basitçe bir toplama işlemi değildir.

ama ne olursa olsun 0.05'ten büyük olacağı kuşkusuzdur). Bu ise "istatistiksel olarak önemli" kabul edilen maksimum yanılma düzeyinden çok fazladır. Yani toplam hata payı suni olarak artmıştır.

Buna önemlilik enflasyonu (*significance inflation*) adı verilebilir. Bu nedenle ikiden çok grubun ortalamalarının karşılaştırılmasında t testinin kullanılması son derece hatalıdır.

Sözüne güvenilir ve zamanını iyi kullanan bir işadamı size "Yarın saat 10'daki randevuya en az %95 olasılıkla gelirim" dediği zaman gerçekten de geleceğini bilirsiniz. Buradaki "en az %95" olasılığını, 0.05'ten küçük p değeri ile eşdeğer olarak düşünün. Bu işadamı, yoğun bir randevu trafiginin olduğu bir gün, sizden başka beş kişiye de randevu verir ve hepsine de "en az %95 olasılıkla gelirim" derse, bazı randevularına yetişememe olasılığı artacağı için, "%5'ten az olasılıkla randevuya gelemem" anlamına gelmesi gereken sözünü tutamamış olur. Çünkü bir önceki görüşmenin uzaması durumunda sonraki randevulara ilişkin sözünü tutamayacaktır. Diğer bir tanımla randevulara gelme olasılıkları birbirlerine bağımlıdır.

İşte ikiden çok grubun karşılaştırılmasında ikişerli karşılaştırmalar yapıp, olasılık hesaplarken, karşılaştırılan grupların ortak olması nedeniyle bu olasılıklar da birbirlerine bağımlıdır. Hesaplanan p değerinin gerçek durumdan daha küçük olmasının nedeni budur.

TEK YÖNLÜ VARYANS ANALİZİ

İkiden çok grubun ortalamaları önce global olarak karşılaştırılmalı ve herhangi bir grubun ortalamasının diğer gruptardan herhangi birininkinden farklı olup olmadığı ortaya konulmalıdır. Varyans analizi bu amaçla kullanılan bir istatistik yöntemdir. Grup ortalamalarının karşılaştırıldığı bir testin adının neden varyans analizi olduğu, test istatistiğinin hesaplanmasında grup ortalamaları ile birlikte grup varyanslarının da çok önemli yer almasına bağlanabilir. Gerçekten de grup varyansları, grup ortalamaların göre fazla büyüğse, yani gruptardaki veriler geniş bir aralığa yayılan, küçük Büyüklü verilerse, ilk bakışta grup ortalamaları birbirlerinden çok farklı gibi görünse bile, varyans analizi ile bunun tersi bir sonuç elde edilebilir.

Varyans analizi sonucunda hangi grup ya da grupların ortalamalarının, diğer grupların hangisi ya da hangilerinden farklı olduğu belli olmaz. Bunu çözümlemek için ikişerli karşılaştırmalar gereklidir, ama t testi yerine çoklu karşılaştırma yöntemleri (*multiple comparison procedures*) adı verilen istatistik yöntemler kullanılmalıdır.

İkiden çok bağımsız örneklem ortalamasının karşılaştırılması tek-yönlü

varyans analizi (*One-way ANalysis Of VAriance: One way ANOVA*) ile yapılır. Tek-yönlü terimi grupları birbirinden ayıran tek özellik olduğu, ya da grupların tek değişkenin değerleri ile ayrıldığı anlamına gelmektedir. Yukarıdaki örnekte "effüzyon nedeni", dört grubu birbirinden ayıran, ya da grupları belirleyen değişkendir. Bu değişkenin değerleri "primer akciğer kanseri, plevra metastazı, parapnömonik effüzyon ve tüberküloz" olduğuna göre, dört bağımsız grubu belirleyen tek değişken "effüzyon nedeni" değişkenidir. Eğer birden çok değişkenle bağımsız gruplar belirlenirse, o zaman yapılacak teste iki-yönlü ya da "faktöriyel" varyans analizi adı verilir. Örneğin hem "effüzyon nedeni", hem de "cinsiyet" grupları arasında ADA düzeylerinin ortalamalarının farklı olup olmadığı test edilecekse, iki-yönlü varyans analizi uygulanmalıdır.

Varyans analizinin mantığı

Örnek 48'deki çalışmada tüberküloz grubunun ortalama ADA düzeyini 120 IU ve tüberküloz grubu da dahil olmak üzere tüm gruppardaki 80 hastaının ortalama ADA düzeyinin ise 65 IU olduğunu varsayılmıştır.

Tüm grubun ortalamasına büyük ortalama (*grand mean*) adı verilir. Tüberküloz grubundaki bir hastada ADA düzeyi 105 IU olsun. Bu hastanın ADA düzeyi büyük ortalamadan, yani 65 IU'den 40 IU daha yüksek, tüberküloz grubu ortalamasından ise 15 IU daha düşüktür. Varyans analizinde tüm hastaların ADA düzeylerinin dahil oldukları grubun ortalamasından ve büyük ortalamadan ne kadar farklı olduğu dikkate alınır ve ADA düzeylerinin hem her grubun kendi içinde ne kadar farklılık gösterdiği (grup içi varyans), hem de grup ortalamalarının birbirlerinden ne kadar farklılık gösterdiği (gruplar arası varyans) bulunur. Gruplar arası varyans ne kadar büyükse, grup ortalamalarının birbirlerinden farklı olma olasılığı o kadar fazladır. Öte yandan grup içi varyans ne kadar büyükse, grup ortalamalarının birbirlerinden farklı olma olasılığı o kadar azdır.

Varyans analizinde her ne kadar grup içi ve gruplar arası varyanslar karşılaştırılıyorsa da, aslında grup ortalamaları karşılaştırılmaktadır. Gruplar arası varyans, grup içi varyanstan büyükse, "grup ortalamaları eşit değildir" sonucuna varılır.

Varyans analizindeki terimler

Varyans analizi elde yapmak için oldukça fazla matematik işlemi gerektirtiği için formülleri verilmeyecektir. Ancak değişik bilgisayar programlarında aynı anlama gelen farklı terimler kullanıldığı için varyans analizi tablosundan ve bazı terimlerden kısaca söz edilecektir.

Şekil 1'de bir varyans analizi tablosu örneği görülmektedir. Tabloda en son sütundaki p değeri <0.05 olduğu için grup ortalamalarının birbirlerinden farklı olduğu anlaşılmaktadır. Bir altta da aynı varyans analiz tablosu SPSS'te kullanılan terimlerle verilmiştir.

Varyans kaynağı	Kareler toplamı	Serbestlik derecesi	Kareler ortalaması	F	p
Gruplar arası	66.150	2	33.075	142.477	.000
Grup içi	9.750	42	.232		
Toplam	75.900	44			

ANOVA					
SELENYUM					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	66.150	2	33.075	142.477	.000
Within Groups	9.750	42	.232		
Total	75.900	44			

Şekil 1
Varyans analizi tablosu örneği.

Varyans analizinde grup ortalamaları arasında fark olup olmadığı araştırıldığı için grupları belirten değişken bağımsız değişken (*independent*) ya da faktör (*factor*), karşılaştırılan ortalamaları belirten değişken bağımlı değişken (*dependent*) olarak adlandırılır. Bağımsız değişken sayısı tekse, varyans analizine tek-yönlü varyans analizi (*one-way analysis of variance*), iki ya da daha çoksa iki-yönlü varyans analizi (*two-way analysis of variance*) adı verilir. Yukarıdaki örnekte "Effüzyon nedeni" bağımsız, "ADA düzeyi" bağımlı değişkendir. Bağımsız değişkenin ayırdığı grup sayısına faktörün düzey sayısının adı verilir. *Level* ya da *treatment* bu anlamda kullanılmaktadır. Her hastanın ADA düzeyinin büyük ortalamadan ne kadar değiştigini belirleyen iki "değişim kaynağı"ndan biri grup içi varyans, hata ya da rezidüel, diğer ise gruplar arası varyans olarak tanımlanır. Çeşitli bilgisayar programlarında, değişim kaynağı, *source of variation* ya da kısaca *source* olarak; grup içi varyans, *variance within groups*, *within groups*, *within*, *error* ya da *residual* olarak ve gruplar arası varyans ise *variance among groups*, *variance between groups*, *among groups*, *between groups* ya da kısaca *between* olarak tanımlanır.

Varyans analizinde kaç grup olursa olsun tek bir F değeri ve buna karşılık gelen tek bir p değeri hesaplanır. Eğer p değeri <0.05 ise, grup ortalamaları eşit değildir. Hangi grup ya da grupların ortalamalarının farklı olduğunu araştırmak için çoklu karşılaştırma yöntemleri uygulanmalıdır (t testi değil!). Eğer varyans analizinde elde edilen p değeri >0.05 ise grup ortalamaları eşittir. Bu durumda çoklu karşılaştırma testlerine gerek kalmaz.

Varyans analizinin varsayımları

İki grup ortalamasının t testi ile karşılaştırılabilmesi için dağılımin normal olması ve varyansların eşit olması koşulunun arandığından söz edilmiştir. İkiinden çok grup ortalamasının varyans analizi ile karşılaştırılabilmesi için sağlanması gereken dört koşul vardır:

1. Her gruptaki değerler kendi içinde normal dağılım göstermelidir. Denek sayısı ne kadar fazlaysa, bu koşulun gerekliliği o kadar azalır.
2. Grup varyansları eşit olmalıdır. Eğer gruplardaki denek sayıları yaklaşık eşitse, varyansların eşit olmaması önemli değildir. Denek sayıları ve varyanslar eşit değilse, ya uygun dönüşümlerle varyansların eşitliği sağlanmalıdır, ya da her gruptan tesadüfi olarak eşit sayıda denek seçilip, denek sayıları eşitlenmelidir. Bazı deneklerin çalışmadan çıkarılması anlamına gelen bu ikinci çözüm yoluna harcanan emeklerin boş gitmesi nedeniyle nadiren başvurulur.
3. Varyans analizinin bir varsayıımı da varyansların ortalamalardan bağımsızlığını olmalıdır. Ortalama yüksek olan grubun varyansı yüksek, ortalama düşük olan grubun varyansı düşükse, varyanslar ortalamalardan bağımsız değildir. Bu durumda da verilere uygun dönüşümler uygulanarak, varyansların ortalamadan bağımsızlığı sağlanmalıdır.
4. Son koşul "grupların bağımsızlığı"dır. Örneğin hastalardan dört farklı zamanda alınan kan örneklerinde yapılan bir çalışmada dört veri grubu aynı hastalardan elde edildiği için gruplar bağımlıdır. Bu durumda varyans analizinin bir uyarlaması olan tekrarlı ölçümler varyans analizi uygulanmalıdır.

Gruplardaki denek sayıları fazla ve aşağı yukarı eşitse, grupların bağımsızlığı olması koşulu ile, normal dağılım, varyansların eşitliği ve varyansların ortalamadan bağımsızlığı koşullarından vazgeçilebilir. Aksi takdirde, ya dönüştüm uygulayarak, koşullar sağlanmalı, ya da tek yönlü varyans analizinin nonparametrik karşılığı olan Kruskal-Wallis analizi uygulanmalıdır.

Varyanslar eşit olmazsa ne olur?

Varyans analizi yapılmabilmesi için gerekli olan koşullardan birinin grup varyanslarının eşitliği olduğu belirtilmişti. Çünkü varyans analizinde grup

ortalamalarının birbirlerinden ne kadar farklı olduğu kadar, her grubun içindeki değerlerin grup ortalamasından ne kadar uzak ya da yakın olduğu da önemlidir. Gruplar içindeki değişkenlik, yani varyans ne kadar fazlaysa, grup ortalamaları arasındaki farkın önemli olması olasılığı o kadar azalır.

ÖRNEK 49

Siroz, kronik aktif hepatit (KAH) ve karaciğer dışı hastalıklara bağlı asitte asit sıvısında mangan (Mn) düzeyi ölçülüyor. Üç grubun Mn ortalamaları arasında fark olup olmadığı varyans analizi ile inceleniyor. Şekil 2'de bu çalışmaya ait varyans analizi tablosu görülmektedir.

Descriptives								
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 Siroz	14	1178.57	1014.65	271.18	592.73	1764.41	100	3500
2 KAH	11	1227.27	1137.62	343.01	463.01	1991.54	100	3500
3 Non-hepatik	12	480.00	663.85	191.64	58.21	901.79	50	2300
Total	37	966.49	991.50	163.00	635.90	1297.07	50	3500

ANOVA					
MANGAN1	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4217854	2	2108927	2.300	.116
Within Groups	3.1E+07	34	916852.6		
Total	3.5E+07	36			

Şekil 2
Örnek 49'daki çalışma gruplarındaki serum Mn düzeylerinin tanımlayıcı istatistikleri ve varyans analizi tablosu.

Şekil 2'deki tablodan anlaşılacağı gibi gruppardaki Mn düzeyi ortalaması 1179, 1227 ve 480'dir. Sirozlu ve KAH'lı hastalarda asit Mn düzeyinin non-hepatik asitli hastalardan daha yüksek olduğu görülmektedir. Tüm gruplar birleştirildiğinde ise Mn ortalamasının 966 olduğu görülmektedir.

Varyans analizinde bir anlamda, her grubun ortalamasının 881'den farklı

olup olmadığı test edilmektedir. Tabloda ilk satırın sonunda verilen *Sig.* değeri, *p* değerine karşılık gelmektedir ve sürpriz olarak (!) 0.05'ten büyütür. Yani grup ortalamaları arasındaki fark önemli bulunmamıştır.

Bunun nedeni grup içi değişkenliğin çok fazla olmasıdır. Üç grupta da değişkenliğin ölçütü olan standart sapmanın ortalamaya göre çok geniş olduğu görülmektedir. Bunun nedeni, bu gruptarda ortalamadan çok küçük ya da çok büyük değerlerin olmasıdır.

Non-hepatik asitli gruptaki ortalamadan çok küçük ve çok büyük değerleri ortalamaya yaklaşacak şekilde değiştirek, Şekil 3'deki tabloyu elde ettigimizde, bu grubun standart sapmasının azaldığını ve ortalama değerler değişimenesine rağmen *p* değerinin de biraz azaldığını görüyoruz, ama hala üç grubun ortalamaları birbirinden farklı değil.

Descriptives									
				95% Confidence Interval for Mean					
		N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
1 Siroz	14	1178.57	1014.65	271.18	592.73	1764.41	100	3500	
2 KAH	11	1227.27	1137.62	343.01	463.01	1991.54	100	3500	
3 Non-hepatik	12	480.00	304.75	87.97	286.37	673.63	250	1200	
Total	37	966.49	936.38	153.94	654.28	1278.69	100	3500	

Şekil 3
Şekil 2'deki tabloda üçüncü grubun standart sapması küçüldüğünde meydana gelen değişiklikler.

ANOVA					
MANGAN2					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4217854	2	2108927	2.622	.087
Within Groups	2.7E+07	34	804323.2		
Total	3.2E+07	36			

Bu kez KAH grubundaki aşırı değerleri de ortalamaya yaklaşırarak, bu grubun da standart sapmasını küçültelim (Şekil 4). Grup ortalamaları değişmediği halde, bu kez ortalamalar arasındaki farkın önemini olduğunu anlıyoruz, çünkü *p* değeri <0.05 oldu.

Descriptives								
MANGAN3	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 Siroz	14	1178.57	1014.65	271.18	592.73	1764.41	100	3500
2 KAH	11	1227.27	575.92	173.85	840.37	1614.18	600	2100
3 Non-hepatik	12	480.00	304.75	87.97	286.37	673.63	250	1200
Total	37	966.49	780.67	128.34	706.20	1226.77	100	3500

ANOVA					
MANGAN3	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4217854	2	2108927	4 046	.027
Within Groups	1.8E+07	34	521235.0		
Total	2.2E+07	36			

Şekil 4

Şekil 3'deki tabloda ikinci grubun standart sapması küçüldüğünde meydana gelen değişiklikler.

İKİ YÖNLÜ VARYANS ANALİZİ

Plevral sıvıdaki ADA düzeyine etki eden durumlardan birinin effüzyonun nedeni (primer akciğer kanseri, plevra metastazı, ...) olduğu varsayımları yukarıda söz edilen tek-yönlü varyans analizi ile incelenir. Acaba effüzyonun nedeninden başka, ADA düzeyine etki eden etkenler de (örneğin cinsiyet) var mıdır? Bu soruyu yanıtlamak için iki-yönlü varyans analizi yapılmalıdır. Çünkü ADA düzeyine etki etmesi beklenen "effüzyonun nedeni" faktöründen başka "cinsiyet" faktörü de birlikte ele alınacaktır. Faktör sayısı üç, dört ya da daha fazla da olabilir. Bu durumda da yapılacak olan analiz iki-yönlüdür. Yani iki-yönlü varyans analizi terimindeki "iki", faktör sayısı değildir.

Dört farklı hastalık grubunun her birinde erkek ve kadın olarak iki alt grup olacağı için toplam 8 grubun ortalamalarının karşılaştırılacağı iki-yönlü varyans analizinde şu sorulara yanıt bulunabilir:

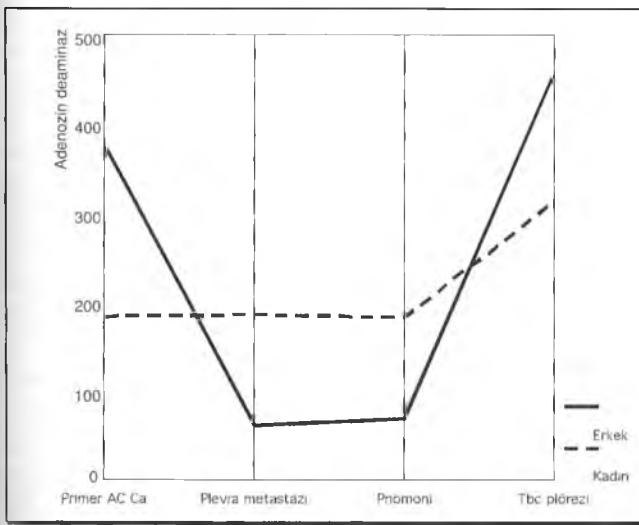
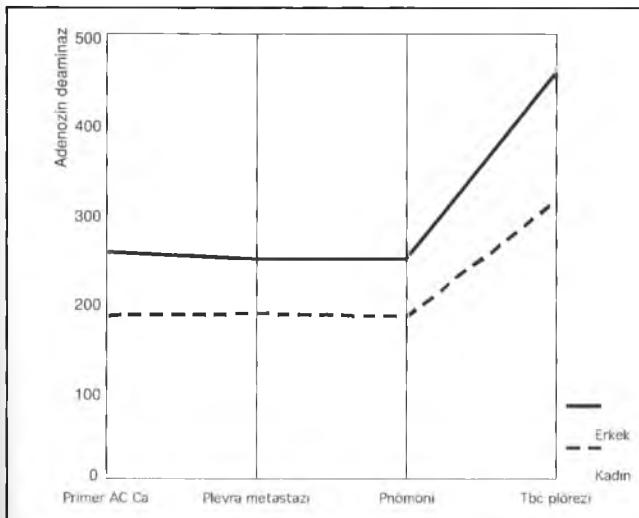
1. Farklı hastalık gruplarında plevral sıvıdaki ADA ortalamaları farklı mıdır?
2. Erkek ve kadınlarda plevral sıvı ADA ortalamaları farklı mıdır?
3. ADA ortalamaları arasında yalnızca hastalık grubu ya da cinsiyetle açıklanmayan, ama bu iki faktörün kombinasyonları ile açıklanan farklılıklar

var mıdır? Örneğin primer akciğer kanseri grubunda erkeklerin ADA ortalamaları kadınlardan yüksek iken, tüberküloz grubunda kadınların ADA düzeyi erkeklerden yüksekse, iki faktör arasında etkileşim (*interaction*) var demektir.

Etkileşim konusunu aydınlatabilmek için Şekil 5'deki grafikleri inceleyelim. Her iki grafikte de yatay eksende "effüzyon nedeni"nin dört farklı düzeyi ve dikey eksende ise ADA düzeyi olmak üzere erkekler için 4 ve kadınlar için 4 olmak üzere toplam 8 grubun ortalamaları noktalanmış ve iki cinsiyetin grup ortalamaları birleştirilerek bir grafik elde edilmiştir.

Şekil 5'deki birinci grafikten tüberküloz grubunda diğer hastalık gruplarına göre ADA düzeyinin daha yüksek olduğu görülmektedir. Ayrıca iki eğri yaklaşık paralel olduğu için erkeklerde ADA ortalamasının tüm hastalık gruplarında kadınlardan daha yüksek olduğu da belirgindir. Bu örnekte "effüzyon nedeni" ile "cinsiyet" arasında etkileşim yoktur.

İkinci grafikte ise ADA ortalamasının primer akciğer kanseri ve tüberkülozda erkeklerde, plevra metastazı ve parapnömonik effüzyonda ise kadınlar da daha yüksek olduğu görülmektedir. Bu nedenle iki eğri paralel değildir. Yani "effüzyon nedeni" ile "cinsiyet" arasında etkileşim vardır.



Şekil 5

Farklı nedenlere bağlı plevral effüzyonlarında ve farklı cinsiyetlerde adenozin deaminaz (ADA) ortalamaları. Yukarıdaki grafikte eğriler paralel olduğu için cinsiyet ile effüzyon nedeni arasında etkileşim yoktur. Aşağıdaki grafikte ise eğriler kesiştiği için cinsiyet ile effüzyon nedeni arasında etkileşim vardır.

İki faktör arasında etkileşim varsa, faktörlerin tek tek etkilerinden söz etmek zorlaşır. Etkileşimin olmadığı birinci grafikteki örnekte hangi cinsiyet olursa olsun tüberkülozda ve hangi hastalık olursa olsun erkeklerde ADA düzeyi daha yüksek olarak görülmektedir. İki faktör arasında etkileşimin olduğu

ikinci grafikte ise erkeklerde mi, kadınarda mı ya da tüberkülozda mı, primer akciğer kanserinde mi ADA düzeyinin yüksek olduğu söylenemez. Bu nedenle faktörler arasında etkileşim varsa, iki-yönlü varyans analizi yapılamaz.

İki-yönlü varyans analizinin mantığı da tek-yönlü varyans analizi gibidir. Her denek değerinin ait olduğu grupların ortalaması ve büyük ortalamadan ne kadar farklı olduğuna dayanan hesaplamalarla faktörlerin tek tek etkileri ve aralarındaki etkileşimler bulunur. Örneğin tüberkülozlu bir kadın hastanın ADA düzeyinin, hem cinsiyeti ne olursa olsun tüm tüberkülozlu hastaların, hem de hastalığı ne olursa olsun tüm kadınların ortalama ADA düzeyinden ne kadar farklı olduğu önemlidir.

Faktör sayısı arttıkça, hem her altgruba düşen denek sayısı azalacağı için, hem de ortaya çıkacak etkileşimler nedeniyle iki-yönlü varyans analizinin yorumlanması güçleşir.

ÇOKLU KARŞILAŞTIRMA YÖNTEMLERİ

Plevral sıvıda ADA çalışmasındaki dört grubun hepsinin diğerleriyle karşılaştırılması amaçlanmışsa, önce varyans analizi uygulanması gereklidir. Varyans analizi ile hesaplanan p değeri <0.05 ise, grupların ADA ortalamalarının eşit olmadığı anlaşılır.

Peki hangi grup ya da grupların ADA ortalamaları diğerlerinden farklıdır? Bu soruyu yanıtlamak için t testi kullanılmasının daha önce söz edilen sakıncaları nedeniyle başka yöntemler uygulamak gereklidir. Bu yöntemler genel olarak çoklu karşılaştırma yöntemleri (*multiple-comparison procedures*) olarak adlandırılır. Bu yöntemlerle yalnızca ikişerli karşılaştırma yapılmaz; bir grubun ortalaması diğer iki ya da daha çok grubun ortalamalarının ortalaması ile de karşılaştırılabilir. Örneğin biri placebo ve diğerleri ise bir ilacın üç farklı dozu kullanıldığı dört grup olduğunu düşünelim. Her grubun ortalaması, diğerleri ile karşılaştırılabileceği gibi, ilaç verilen üç grup birleştirilerek elde edilen ortalama, placebo grubu ile de karşılaştırılabilir.

Yukarıdaki çalışmada asıl amacın, değişik nedenlere bağlı plevral effüzyonlarda ADA düzeylerinin karşılaştırılması değil de, (a) akciğer kanseri ve akciğer dışı kanserlere bağlı effüzyonlarda ADA düzeylerinin karşılaştırılması ve (b) tüberküloz ve parapnömonik effüzyonlarda ADA düzeylerinin karşılaştırılması olduğunu düşünelim. Yani örneğin akciğer kanseri ile parapnömonik effüzyon gruplarındaki ADA düzeylerinin karşılaştırılması ile ilgilenilmiyor olsun.

Yukarıda (a) ve (b) ile gösterilen iki karşılaştırmada çakışan hasta grupları yoktur. (a)'da adı geçen iki grup (b)'deki, (b)'de adı geçen iki grup (a)'da

karşılaştırmada yer almamaktadır. Bu tür ikili karşılaştırmalara ortogonal karşılaştırmalar adı verilir. Ortogonal karşılaştırmalarda, varyans analizi yapılmasına gerek olmayabilir. Çünkü ortogonal karşılaştırmalarda çoklu t testi kullanılması ile bölümün başında söz ettiğimiz toplam önemlilikteki suni artış olmaz. Bu durumda bu iki karşılaştırma Student t testi ile yapılabilir.

Şimdi de yukarıdaki çalışmanın amacının tüberküloza bağlı effüzyonlarda ADA düzeyinin diğer klinik durumlardan farklı olup olmadığına saptanması olduğunu düşünelim. Bu durumda tüberküloz-akciğer kanseri, tüberküloz-plevra metastazı ve tüberküloz-parapnömonik effüzyon olmak üzere üç ayrı ikili karşılaştırma yapılmalıdır. Bu karşılaştırmalar ortogonal değildir, çünkü hepsinde de tüberküloz grubuna ait veriler kullanılmaktadır. Bu durumda bu karşılaştırmalar ortogonal değildir ve Student t testi kullanılması ile önemlilik enflasyonu ortaya çıkabilir.

Önceden planlanmış (*a priori*) karşılaştırmalar

Hangi grupların ortalamalarının ikişerli olarak karşılaştırılacağı önceden belirlenmişse, yani ikişerli karşılaştırmaların hepsi ile değil, bazıları ile ilgileniliyorsa, varyans analizi yapılmadan grup ortalamalarının arasında fark olup olmadığı test edilebilir. Ancak bu durumda yanılma payındaki artmayı ortadan kaldırmak için her ikişerli karşılaştırmanın yanılma düzeyi 0.05'i ikili karşılaştırma sayısına bölerek aşağı çekilir.

Bölümün başındaki tabloda p değerlerinin aşağıdaki gibi olduğunu düşünelim.

- | | |
|---|----------|
| 1. Primer akciğer kanseri ile plevra metastazı | p=0.20 |
| 2. Primer akciğer kanseri ile parapnömonik effüzyon | p=0.04 |
| 3. Primer akciğer kanseri ile tüberküloz | p=0.0001 |
| 4. Plevra metastazı ile parapnömonik effüzyon | p=0.03 |
| 5. Plevra metastazı ile tüberküloz | p=0.0001 |
| 6. Parapnömonik effüzyon ile tüberküloz | p=0.03 |

Buradaki altı ikili karşılaştırmanın hepsiyle ilgilenilmediğini, yalnızca tüberküloz grubu ile diğer gruplar arasında fark olup olmadığını araştırdığımı düşünürsek, 3, 5 ve 6 numaralı ikili karşılaştırmaların (t testinden elde edilmiş olan) p değerlerine bakalım. Bu p değerlerinin hepsi de 0.05'ten küçüktür. Ancak üç tane ortogonal olmayan (tümünde de tüberküloz grubu var) ikili karşılaştırma yapılacak için yanılma düzeyi karşılaştırma sayısı olan 3'e bölünmelidir.

Böylece hesaplanan p değerinin 0.05'ten değil, $0.05/3$, yani ancak 0.017'den küçük olması durumunda, iki grubun ortalamalarının istatistiksel olarak birbirinden farklı olduğu söylenebilir. Böylece önemlilik enflasyonu

sorunu çözülebilir. Bu durumda tüberküloz ile parapnömonik effüzyon gruplarının ADA ortalamalarının farklı olmadığı anlaşılır. Çünkü Student t testi ile hesaplanan p değeri 0.05'ten küçük olmasına karşın, 0.017'den büyktür.

Tüm ikili karşılaştırmalar t testi ile yapılrsa, bu durumda yanlışma düzeyi 0.05/6, yani 0.008 olarak seçilmelidir. Bu durumda 2, 4 ve 6 numaralı karşılaştırmaların p değerleri 0.05'ten küçük olduğu halde, 0.008'den büyük oldukları için bu grupların ortalamalarının farklı olmadığı anlaşılır.

Bonferroni t yöntemi (ya da Dunn çoklu karşılaştırma yöntemi) de, yanlışma düzeylerinin aşağıya çekilmesine eşdeğer bir yöntemdir.

Sonradan yapılan (*posteriori: post hoc*) karşılaştırmalar

Varyans analizi ile grup ortalamalarının arasında fark olmadığı saptanırsa, ikili karşılaştırma yapmaya gerek kalmadığı açıklır. Eğer varyans analizi ile bulunan p değeri <0.05 ise, ikili karşılaştırmaların yapılması gereklidir. Varyans analizi sonrası yapılan çoklu karşılaştırma yöntemlerine *post hoc* yöntemler adı verilir. Çeşitli *post hoc* çoklu karşılaştırma yöntemleri vardır.

Tukey HSD yöntemi

Bazı *post hoc* çoklu karşılaştırma yöntemleri ile grup ortalamaları arasında önemli fark bulmak oldukça kolaydır, yani yanlışma düzeyi 0.05 olarak seçilebilir, %5'ten daha fazla olasılıkla gerçekte fark olmadığı halde, fark varmış gibi sonuç elde edilir. Tukey testi ise varyans analizi sonrası uygulanan çoklu karşılaştırma yöntemleri içinde en doğru ve en güclü testlerden biridir. Bu nedenle bu teste Tukey'in gerçekten önemli farklar testi (*Tukey's honestly significant difference test - HSD*) denilmektedir.

Tukey testi ile yalnızca ikişerli karşılaştırmalar yapılabilir. Bu testle HSD adı verilen en küçük önemli fark hesaplanır. Eğer iki grubun ortalamaları arasındaki fark, HSD'den fazlaysa, iki grubun ortalamalarının birbirinden önemli derecede farklı olduğu söylenebilir. Örneğin ADA ortalaması tüberküloz grubunda 120, parapnömonik effüzyon grubunda 85 ve akciğer kanseri grubunda 25 ise, ve hesaplanan HSD 40 ise, tüberküloz ve parapnömonik effüzyon grubunun ADA ortalamalarının birbirinden farklı olmadığı, ancak her iki grupta da ADA ortalamasının akciğer kanserinden farklı olduğu sonucu anlaşılır.

Scheffe yöntemi

Scheffe yöntemi, *post hoc* yöntemler içinde en kullanışlı olanıdır. Yalnızca ikişerli karşılaştırmalar değil, grupların birleştirilmiş ortalamaları da karşılaştırılabilir. Örneğin malign hastalıklar ve benign hastalıklarda plevral sıvı ADA ortalamalarının karşılaştırılmak istendiğini varsayıyalım. Bu durumda

“akciğer kanseri” ve “plevra metastazı” grupları birleştirilip “malign grup”, “tüberküloz” ve “parapnömonik effüzyon” grupları birleştirilip “benign grup” olarak adlandırıldığında, bu iki grubun ortalamaları arasında fark olup olmadığı Scheffe yöntemi ile test edilebilir. Ancak ortalamalar arasındaki farkın önemli olarak saptanabilmesi diğer testlere göre (Tukey testi dahil) daha zordur. Tukey testinde olduğu gibi hesaplanan en küçük önemli fark, grup ortalamaları arasında farkla karşılaştırılır.

Newman-Keuls yöntemi

Bu yöntemde gruplar ortalamalarına göre küçükten büyüğe dizilir ve grupların bu sıralamada yakınlık derecelerine göre farklı en küçük önemli farklar hesaplanır. Örneğin ADA ortalaması tüberküloz grubunda 120, parapnömonik effüzyon grubunda 85, plevra metastazında 45 ve akciğer kanserde 25 ise, tüberküloz-parapnömonik effüzyon karşılaştırılmasında kullanılacak olan en küçük önemli fark, tüberküloz-plevra metastazı karşılaştırılmasında kullanılacak olan farktan daha küçük olarak hesaplanır. Örneğin en küçük önemli fark komşu gruplar için 30, arada bir grup bulunan gruplar için 40, arada iki grup bulunan gruplar için 55 olabilir (Şekil 6).

Tüberküloz	Parapnömonik effüzyon	Plevra metastazı	Akciğer kanseri
120	85	45	25
30	30	30	
40		40	
		55	

Şekil 6 Newman-Keuls yönteminde gruplar ortalamalarına göre sıraya dizildiğinde birbirine komşuluğu, yani ortalamaları yakın grupların ortalamaları arasında fark olabilmesi için daha küçük bir en küçük önemli fark yeterli olurken, uzak gruplar için daha büyük bir en küçük önemli fark hesaplanır. Şekilde her grubun adının altında ADA ortalamaları ve bunların altında da en küçük önemli farklar görülmektedir.

Komşu gruplar arasında fark bulunabilmesi, Tukey testine göre daha kolaydır. Arada bir grup bulunan grupların karşılaştırılmasında ise Tukey testi ile genellikle aynı sonucu verir.

Dunnett yöntemi

Yalnızca birden çok grubun tek bir grupta karşılaştırılması amacıyla kullanılır. Örneğin plevral sıvıda ADA çalışmasının amacı tüberkülozda ADA düzeyinin diğer klinik durumlardan farklı olup olmadığı ise, bu yöntem uygundur. Tüberküloz dışı plevral effüzyon nedenlerinin kendi aralarında karşılaştırılmasına izin verilmez. Bu kısıtlama nedeniyle sık kullanılmaz, ama Tukey ve Scheffe testlerine göre grup ortalamaları arasındaki farkın önemli olarak bulunması daha kolaydır.

VARYANS ANALİZİNDE DİĞER DÜZENLER VE YÖNTEMLER

Bağımsız gruplarla klinik çalışma yaparken, gruplardaki deneklerin benzerliği ya da farklılığının kritik öneme sahip olduğu anlatılmıştır. Bu sorunun çözümünde kullanılan çeşitli çalışma düzenleri vardır. Bu düzenlerde amaç grupların istenilen özellikler açısından benzerliğinin sağlanmasıdır.

Randomize blok düzeni

Klinik çalışmalarında hastaların özellikle yaş, cinsiyet, vücut ağırlığı, hastalıkın şiddetini vb. özellikler açısından benzer gruplar oluşturmaları çok önemlidir. Antihipertansif bir ilacın etkinliğinin araştırıldığı bir çalışmada, ilacin iki farklı dozunun plaseboyla karşılaştırıldığını düşünelim. Plasebo verilen gruptaki hastaların diğer iki gruba göre tesadüfen daha hafif hipertansiyonlu olmaları, ilacın etkisiz ya da olduğundan daha az etkili bulunması gibi bir hata yolu açabilir. Bu sorunla karşılaşmak için kullanılacak yöntemlerden biri randomize blok düzeni (*randomised block design*) kurtmaktadır.

Bu çalışma düzeneinde önce hastalar hipertansiyonun şiddetine göre "hafif", "orta" ve "ağır" olarak bloklara bölünür. Daha sonra her bloktaki hastaların, plasebo ya da ilaç gruplarından hangisine alınacağı tesadüfi olarak seçilir. Tablo 1'de hipertansiyonun şiddetine göre bloklara ayrılmış bir randomize blok düzeni örneği görülmektedir.

Tablo 1 Randomize bloklu çalışma düzeni örneği.

Hipertansiyonun şiddeti	Plasebo	İlaç (10 mg)	İlaç (20 mg)
Hafif	1	2	3
Orta	4	5	6
Ağır	7	8	9

Her blokta üç hasta bulunduğuunu varsayıyalım. Hafif hipertansiyonlu üç hastadan biri (1 numaralı hasta) palsebo, biri (2 numaralı hasta) düşük doz ilaç, diğer de (3 numaralı hasta) yüksek doz ilaç grubuna alınacaktır. Bir blokta üç yerine 15 hasta olsaydı, bu kez tedavi gruplarına her bloktan beşer hasta alınamaktu. Sonuç olarak tedavi gruplarındaki hastalar, hipertansiyonun şiddetine göre benzer olarak seçilmişlerdir.

Latin kare düzeni

Bu kez ilaçın etkinliğinin değişik yaş gruplarında farklı olduğunu düşünelim. Örneğin ilaç yaşlıarda kan basincının düşürülmesinde gençlere göre daha etkili ise ve tesadüfen palsebo grubunda yaşlılar, ilaç gruplarında ise gençler çoğunluktaysa, yine ilaçın olduğundan daha az etkili olduğu gibi hatalı bir sonuçla karşılaşmak olasıdır.

İki değişken açısından grupların benzer olmasının sağlanması için uygun çalışma düzenlerinden biri Latin kare düzeni'dir (*Latin square design*).

Bu düzende hastalar, iki etken açısından bloklara bölünür ve her bloktaki hastalar tedavi gruplarından birine alınır. Tablo 2'de yaş ve hipertansiyonun şiddetine göre bloklara ayrılmış bir Latin kare düzeni tablosu örneği görülmektedir. Örneğin yaşı 20-39 arasında olan ve orta derecede hipertansiyonu olan hastalara düşük doz ilaç verilecektir.

Tablodan görüldüğü gibi her yaş grubu ve hipertansiyon şiddeti grubu için her tedavi grubu denenmiştir. Latin kare düzeni ile 9 hastaya üç farklı tedavi grubunun iki değişken açısından benzerliği sağlanabilir.

Tablo 2 Latin kare düzenli çalışma örneği.

Hipertansiyonun şiddeti	20-39	Yaş	
Hafif	Palsebo	İlaç (20 mg)	İlaç (10 mg)
Orta	İlaç (10 mg)	Palsebo	İlaç (20 mg)
Ağır	İlaç (20 mg)	İlaç (10 mg)	Palsebo

TEKRARLI ÖLÇÜMLER DÜZENİ

ÖRNEK 50

Hemodiyaliz sırasında kompleman sistemi aktive olur mu? Bu soruya yanıt bulmak amacıyla bir grup hastada hemodiyalizden önce ve sonra kompleman düzeyleri ölçülp, önce ve sonraki ortalamalar “iki bağımlı grup için Student t testi” ile karşılaştırılabilir. Bu örnekte olduğu gibi önce-sonra düzeni olarak adlandırılabilen klinik çalışmalar yapmak sıkılıkla başvurulan bir yöntemdir. Hemodiyaliz sonrası kompleman düzeyinin artmış olarak bulunduğu varsalım. Peki hemodiyaliz seansının kaçinci saatinde kompleman düzeyi artmaya başlamaktadır? Bunu incelemek için yalnız diyaliz seansından önce ve bitiminde (5. saatte) değil, hemodiyaliz başladıkten sonra 1 ve 3. saatlerde de kompleman düzeyi ölçümülerinin yapıldığını varsayılm.

Örnek 50’de her hastadan dört kez ölçüm yapılmış olacaktır, yani dört bağımlı grup ortalaması elde edilecektir. Eğer ikiden çok bağımlı grup varsa, yani çalışmaya alınan hastalarda ikiden çok zamanda ölçüm tekrarlanırsa, ortalamaları t testi ile ikişer ikişer karşılaştırmakla yanlışlık riski artacağı kuşkusuzdur. Öte yandan varyans analizinin varsayımlarından en önemlisi olan grupların bağımsızlığı koşulu da sağlanmamaktadır. İşte bu durumlar için uygulanacak yöntem tekrarlı ölçümler varyans analizi adı verilen özel bir varyans analizi yöntemidir. Bu yöntemde de, varyans analizi tablosu elde edilir. Örnek 50’deki gibi tek bir hasta grubunda ikiden çok kez ölçülen verilerin ortalaması bu yöntemle karşılaştırıldığında, ortalamaların zamanla gösterdiği değişimin istatistiksel önemi hesaplanabilir.

ÖRNEK 51

Örnek 50’deki çalışmada hemodiyaliz sırasında kompleman sisteminin aktive olduğunu bulunmuştur. Acaba kompleman sisteminin aktivasyonunun, diyaliz sıvısında tampon olarak bulunan maddenin asetat ya da bikarbonat olması ile ilişkisi var mıdır? Bu soruya yanıt bulmak amacıyla Örnek 50’deki çalışma düzeni, tampon olarak birinde asetat, diğerinde bikarbonat kullanılan iki grup hastada yinelemişsinizdir.

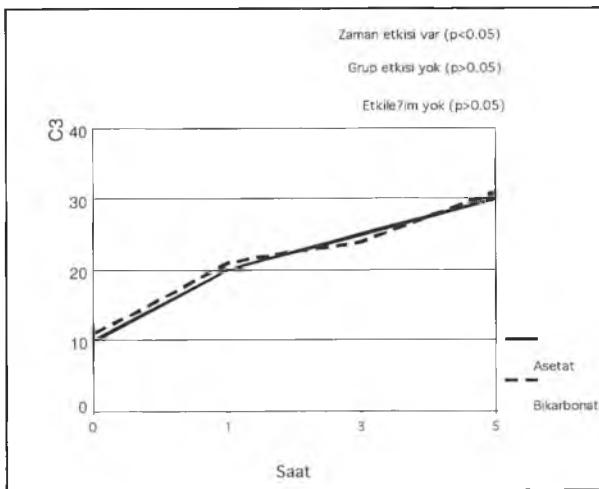
Örnek 51’de kompleman düzeylerinin hem zamanla değişim gösterip göstermediği, hem de bu değişimin farklı tamponların kullanıldığı iki grup arasında farklı olup olmadığı çalışmanın asıl amaçlarıdır. Burada da tekrarlı ölçümler varyans analizi kullanılabilir. Yalnız bu kez zaman dışında tampon ti-

pi de faktör olarak alınacağı için, mantık olarak iki-yönlü varyans analizini andıran bir analiz yapılır. İki-yönlü varyans analizinde olduğu gibi tüm faktörlerin tek bir etkisi ve faktörler arası etkileşimin istatistiksel önemliliği hesaplanır.

Örnek 51'deki çalışmada tekrarlı ölçümler varyans analizi yapıldığında "zaman", "tampon tipi" ve "zaman-tampon tipi etkileşimi"ne ait p değerleri hesaplanacaktır. Eğer "zaman"ın etkisinin istatistiksel önemini temsil eden p değeri 0.05'ten küçükse, kompleman düzeyinin zamanla değişim gösterdiği, "tampon tipi"nin etkisinin istatistiksel önemini temsil eden p değeri 0.05'ten küçükse, farklı tipte tampon kullanıldığında kompleman düzeylerinin farklılık gösterdiği anlaşılır.

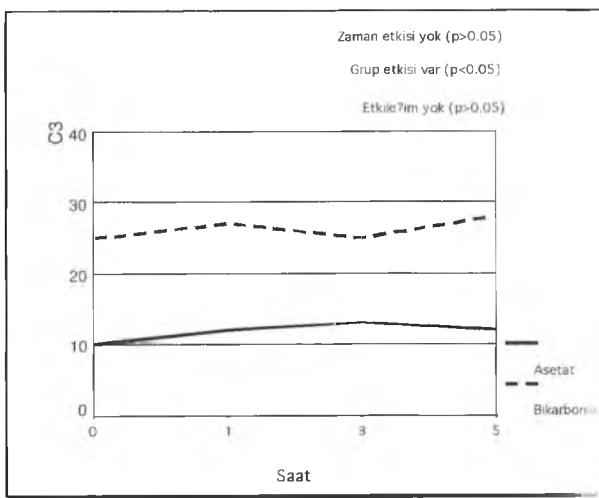
İki faktör arasında etkileşim varsa, yani etkileşime karşılık gelen p değeri 0.05'ten küçükse, iki grupta komplemanın zamanla gösterdiği değişimin aynı olmadığı, yani yatay eksen olarak zaman, dikey eksen olarak kompleman düzeyi olmak üzere iki gruptaki kompleman düzeylerinin grafiği çizildiğinde, eğrilerin paralel olmadığı anlaşılır. İki-yönlü varyans analizinde etkileşim olduğu zaman, grup ortalamaları arasında fark olup olmadığı konusunda yorum yapılamayacağı için, iki-yönlü varyans analizinden vazgeçilmesi gerekliliğinden söz edilmiştir.

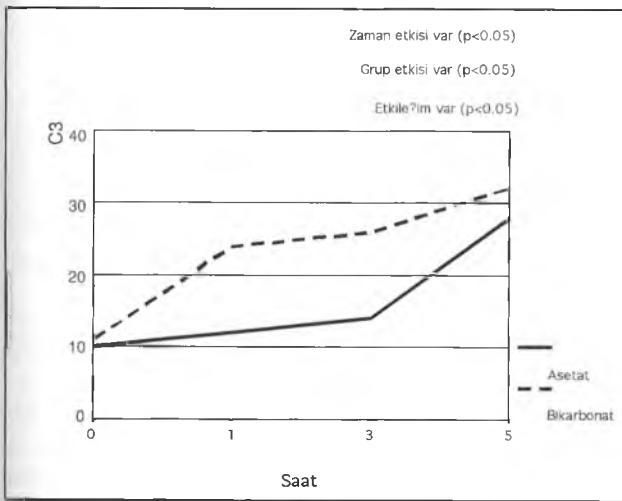
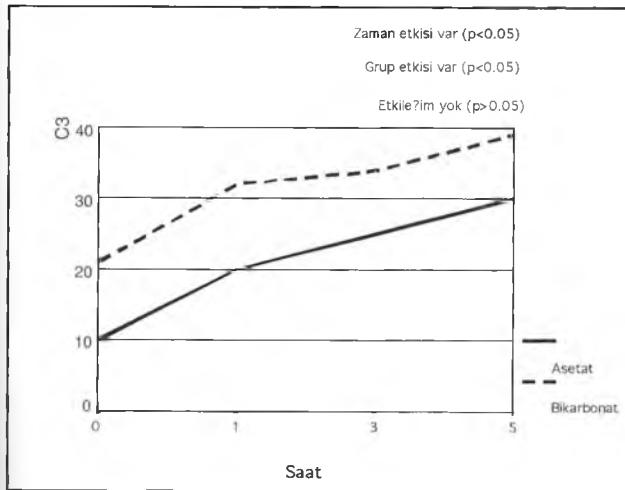
Oysa, tekrarlı ölçümler varyans analizinde etkileşim olduğu halde, sonuçlar yorumlanabilir; hatta bazen etkileşim olması elde edilen bilgiyi artıtabilir (Şekil 7). Tekrarlı ölçümler varyans analizi, ileri istatistik konuları içine girmektedir. Bu nedenle gerektiği zaman mutlaka istatistikte deneyimli uzmanların yardımı istenmelidir.



Şekil 7

Birden çok faktörlü tekrarlı ölçümler varyans analizinde faktör etkileri ve etkileşim. Üstteki şekilde kompleman sisteminin üçüncü elemanı (C_3) her iki tampon tipinde de paralel şekilde zamanla artma göstermektedir. İkinci şekilde C_3 düzeyi her iki grupta da zamanla değişmemektedir, ama iki grup arasında fark vardır.



**Sekil 7 (devamı)**

Üçüncü şekilde C3 düzeyleri hem iki grup arasında farklılık göstermektedir, hem de zamanla artmaktadır. Artış patterni açısından iki grup arasında fark yoktur. En alttaki şekilde ise C3 düzeyleri hem iki grup arasında farklılık göstermektedir, hem de zamanla artmaktadır. Ayrıca artış patternleri açısından da iki grup arasında fark olduğu görülmektedir.

NONPARAMETRİK VARYANS ANALİZİ

Varyans analizi varsayımlarından biri olan gruplardaki değerlerin normal dağılıma uyması her zaman mümkün olmayabilir. Yine de normal dağılımdan aşırı sapmalar olmadıkça, varyans analizi uygulanabilir. Diğer bir varsayımlı olan grup varyanslarının eşitliği koşulundan da, eğer gruplardaki denek sayıları yaklaşık eşitse, vazgeçilebilir. Bu durumlar dışında varyans analizinin varsayımları sağlanmadıkça, sonuçlar yaniltıcı olur. Çözüm olarak ya değişkenlere dönüşüm uygulanır, ya da varyans analizinin nonparametrik karşılığı olan yöntemler uygulanır.

Kruskal-Wallis tek yönlü varyans analizi

Tek yönlü varyans analizinin nonparametrik karşılığı Kruskal-Wallis tek yönlü varyans analizi'dir. Diğer nonparametrik yöntemlerde olduğu gibi Kruskal-Wallis analizi ile grupların ortalamaları değil, ortancaları karşılaştırılır. Eğer Kruskal-Wallis testi ile ortancaların eşit olmadığı saptanırsa (yani $p < 0.05$ bulunursa) post hoc çoklu karşılaştırma yöntemi olarak, yanılma düzeyini aşağı çekerek, Mann-Whitney U testi uygulanır. Örneğin plevral sıvıda ADA çalışmasında, varyans analizi varsayımları sağlanamıyorsa, Kruskal-Wallis testi uygulanabilir. Eğer grup ortancaları farklı bulunursa, toplam altı ikişerli karşılaştırma olduğu için yanılma düzeyi $0.05 / 6 = 0.008$ olmak koşulu ile ikişerli karşılaştırmalar Mann-Whitney U testi ile yapılmalıdır.

İki-yönlü varyans analizinin genel kabul görmüş nonparametrik karşılığı yoktur. Bu nedenle varsayımlar sağlanamadığı takdirde, ancak verilere dönüşüm uygulayarak, iki-yönlü varyans analizi yapılabilir.

Friedman iki-yönlü varyans analizi

Tekrarlı ölçümler varyans analizinin nonparametrik karşılığı, Friedman iki-yönlü varyans analizi'dir. Eğer Friedman testi ile ortancaların eşit olmadığı saptanırsa (yani $p < 0.05$ bulunursa) post hoc çoklu karşılaştırma yöntemi olarak, yanılma düzeyini aşağı çekerek, Wilcoxon testi uygulanır.

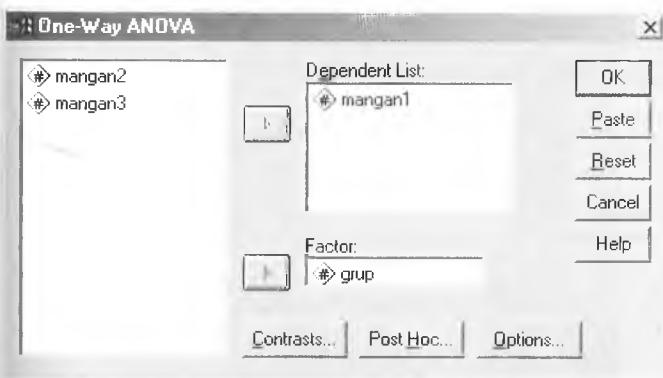
İki-yönlü varyans analizinde olduğu gibi, birden çok faktörlü tekrarlı ölçümler varyans analizinin de nonparametrik karşılığı yoktur. Bu nedenle parametrik test varsayımları sağlanamazsa, verilere dönüşüm uygulayarak, tek riarlı ölçümler varyans analizi yapılabilir.

SPSS'TE VARYANS ANALİZİ NASIL YAPILIR?

Tek yönlü varyans analizi

Analyze >> Compare Means >> One-Way ANOVA

Yukarıdaki seçimle One-Way ANOVA diyalog kutusu açılır (Şekil 8). Sol-daki listede yalnız sayısal değişkenler vardır. Factor kutusuna grupları birbirinden ayıran, yani bağımsız değişken seçilir. Örneğin "hastalık grubu" gibi bir değişken seçilebilir.



Şekil 8
SPSS'te "One-Way ANOVA" diyalog kutusu.

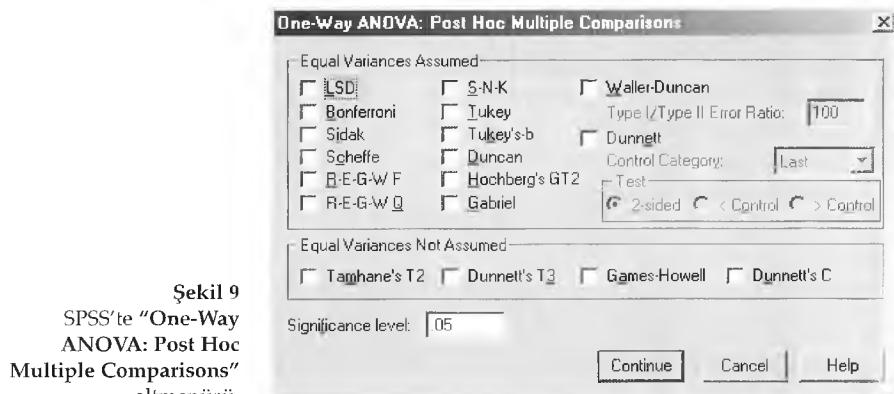
Bağımsız değişken harflerle değil, rakamlarla girilmiştir. Yani sayısal değişken tipinde olmalıdır. Örneğin "hastalık grubu" değişkenini "NOR, SİROZ, KAH" olarak girmişseniz, zaten One-Way ANOVA diyalog kutusundaki değişkenler listesinde de bulamazsınız. Bu nedenle hastalık gruplarını "1, 2, 3" gibi rakamlarla kodlamış olmalısınız.

Ortalamaları karşılaştırılacak değişkenleri ise *Dependent List* kısımına seçiniz. Buradaki bağımlı kavramı, "bağımlı gruplar" ya da "bağımlı örneklem" kavramından farklıdır. Bağımsız değişken olan faktör düzeylerine bağımlılık anlamındadır. Bağımlı değişken olarak birden çok değişken seçebilirsiniz. Her birisi için ayrı varyans analizi tablosu verilecektir.

Options seçeneği ile bağımlı değişkenin her grup için tanımlayıcı istatistiklerini ve %95 güven sınırlarını elde edebilirsiniz. Yine bu menüde Levene varyansların homojenliği testi de vardır. Aslında tüm varyans analizlerinde varyansların homojenliğinin test edilmesi gereklidir. Grup varyansları homojen değilse bile, SPSS varyans analizi tablosunu ve istatistik sonuçlarını verir, ama

bu durumda bu test sonuçları dikkatle kullanılmalıdır.

Post hoc seçeneği ile varyans analizi sonrası ikişerli karşılaştırma testlerinden bir ya da bir kaçının yapılmasını seçebilirsiniz (Şekil 9). *Post hoc* terimi “iş bittiğten sonra” anlamına gelmektedir. Eğer, henüz varyans analizi yapmadan önce belirlenmiş grupların ortalamaları arasında fark olup olmadığı asıl ilgilenilen konu ise, yapılacak olan işleme “*a priori*” yani “*işten önce*” yapılan işlemler adı verilir. *A priori* karşılaştırma yöntemlerine *Contrasts* adı verilen menü ile ulaşılır. Ama bu yöntemler oldukça karmaşıktır ve iyi planlanmış çalışma düzenlerinde uygulanabilir, ayrıca ileri derecede istatistik yorumlama deneyimi gerektirir. Bu nedenle pratikte kullanılan testler post hoc karşılaştırma yöntemleridir.



Şekil 9
SPSS'te “One-Way ANOVA: Post Hoc Multiple Comparisons” altmenüsü.

SPSS'te verilen *Post hoc* seçenekler arasında en basit *Least-significance difference* ya da en küçük önemli fark yöntemidir. Aslında, bu test ikişerli t testi yapmaya eşdeğerdir ve bu nedenle uygulanması hatalıdır. Çoklu karşılaştırma yöntemleri arasında iki grup ortalamaları arasında en kolay önemli fark bulunan yöntemdir. Hatta varyans analizi ile grup ortalamaları arasında önemli fark bulunmasa bile, bu yöntemle ortalamaları birbirinden farklı olan gruplar bulunabilir.

Örneğin Şekil 10'da varyans analizi ile üç grup ortalamaları arasında fark olmadığı gösterilmiş olmasına ($p=0.076$) ve uygulanan diğer post hoc testlerle de bu durumun doğrulanmasına karşın, en küçük önemli fark yöntemi ile kontrol grubu ile kemik metastazı grubunun Ca ortalamalarının farklı olduğu gibi bir sonuç çıkmıştır ($p=0.041$). Aslında bu bir artefakttır. Bu örnekte ol duğrı gibi sıklıkla hatalı sonuçlara neden olduğu için en küçük önemli fark yöntemi kullanılmamalıdır.

Descriptives			
CA	N	Mean	Std. Deviation
1 Kontrol	20	8.630	.890
2 Myeloma	20	10.105	2.834
3 Kemik metastaz?	20	10.230	2.964
Total	60	9.655	2.492

ANOVA					
MANGAN1	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4217854	2	2108927	2.300	.116
Within Groups	3.1E+07	34	916852.6		
Total	3.5E+07	36			

Multiple Comparisons					
Dependent Variable: CA			Mean Difference (I-J)		
	(I) GRUP2	(J) GRUP2			
Tukey HSD	1 Kontrol	2 Myeloma	-1.475	.141	
		3 Kemik metastaz?	-1.600	.101	
	2 Myeloma	1 Kontrol	1.475	.141	
		3 Kemik metastaz?	.125	.985	
Scheffe	1 Kontrol	2 Myeloma	-1.475	.166	
		3 Kemik metastaz?	-1.600	.122	
	2 Myeloma	1 Kontrol	1.475	.166	
		3 Kemik metastaz?	.125	.987	
LSD	1 Kontrol	2 Myeloma	-1.475	.059	
		3 Kemik metastaz?	-1.600	.041	
	2 Myeloma	1 Kontrol	1.475	.059	
		3 Kemik metastaz?	.125	.871	
Bonferroni	1 Kontrol	2 Myeloma	-1.475	.178	
		3 Kemik metastaz?	-1.600	.124	
	2 Myeloma	1 Kontrol	1.475	.178	
		3 Kemik metastaz?	.125	1.000	

Şekil 10
 Tek yönlü
 varyans analizi
 ile en küçük
 önemli fark
 yöneminin
 çelişkili
 sonuçları.

Diger post hoc çoklu karşılaştırma yöntemleri arasında yer alan Bonferroni testi kısmen modifiye edilmiş en küçük önemli fark testidir. Duncan, Student-Newman-Keuls ve Tukey testleri ise genellikle varyans analizi ile uyumlu sonuç verirler.

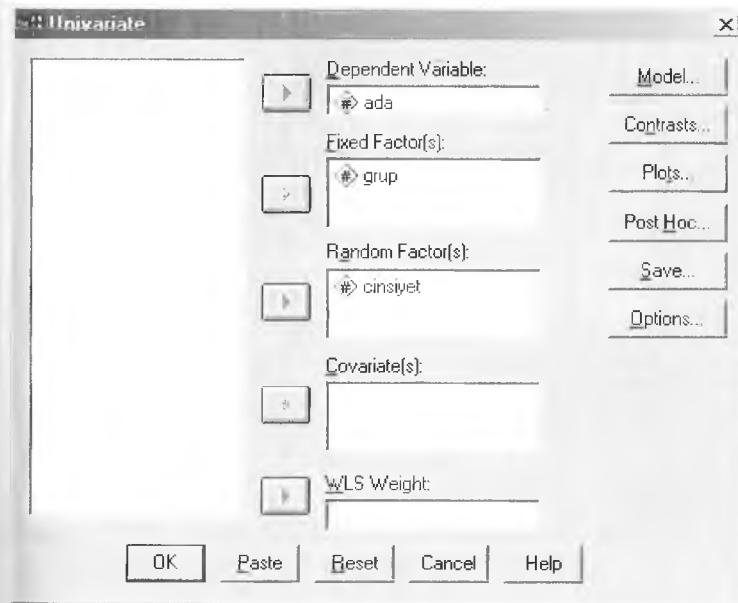
Post hoc testler içinde en küçük önemli fark yöntemi ile gerçekte önemli olmayan farklıların bile önemli olduğu gibi bir sonuç elde edilebilirken, spektrumun diğer ucunda bulunan Scheffé testi ile, varyans analizinde fark olduğu saptansa bile, ikişerli karşılaştırmalarda fark bulunamayabilir. Scheffé testi aslında, ikişerli karşılaştırmalar için tercih edilecek bir yöntem değildir. Gruplar ikişerli olarak değil de, grupların bir kısmı birleşik olarak başka grup ya da gruplarla karşılaşılacaksa kullanılır. Örneğin bir hastalığın tedavisinde A ilaçının iki farklı dozu (5 mg ve 10 mg), B ilaçının iki farklı dozu (5 mg, 10 mg) ve plasebonun kullanıldığı bir çalışmada, 5 grup olduğu için varyans analizi yapılmalıdır. Eğer dozuna bakılmaksızın A ve B ilaçlarının plasebo ile karşılaştırılması istenirse, A ve B ilaçlarının verildiği hastalar A ve B diye iki grupta toplanarak, plasebo ile karşılaştırılabilir. Scheffé testi dışındaki post hoc testler ile yalnız ikişerli gruplar karşılaştırılabilir. SPSS'te Scheffé testinin bu amaçlı kullanımı bulunmamaktadır. Bu amaçla *Contrasts* seçenekinden ortogonal kontrastlar testi yapılmalıdır.

SPSS'te yer alan post-hoc testlerden Tamhane T2, Dunnett's T3, Games-Howell ve Dunnett's C testleri varyansların homojenliği varsayıminın gereklili olmadığı testlerdir.

İki yönlü varyans analizi

Analyze >> General Linear Model >> Univariate

Yukarıdaki seçimle **Univariate** diyalog kutusu açılır (Şekil 11). Burada karşılaştırılacak ortalamaları belirleyen değişken, yani bağımlı değişkeni *Dependent Variable* kutusuna, grupları belirleyen değişkenler, yani faktör ya da bağımsız değişkenleri *Fixed Factor(s)* ya da *Random Factor(s)* kutusuna seçilmelidir (çalışmanın tasarımını gereği önceden planlanmış grupları belirleyen değişkenleri *Fixed Factor(s)* kutusuna, önceden planlanmamış ama analiz sırasında kullanacağımız değişkenleri *Random Factor(s)* kutusuna seçin).



Şekil 11
SPSS'te
"Univariate"
diyalog
kutusu.

Örnek 48'deki çalışmada dört farklı plevral effüzyon grubu ve erkek ve kadınlar ADA ortalamalarının karşılaştırılması amacıyla şunlar yapılmalıdır:

1. Ada değişkenini *Dependent* kutusuna seçin,
2. Grup değişkenini *Fixed Factor(s)* listesine seçin,
3. Cinsiyet değişkenini *Random Factor(s)* listesine seçin,
4. OK'i klikleyin.

Sonuç Şekil 12'de gösterilmiştir. *Sig* sütundaki değerler, *p* değerleridir. "GRUP*CINSIYET" yani iki-yönlü etkileşime karşılık gelen *p* değeri >0.05 (0.766) olduğuna göre cinsiyetle effüzyon nedeni arasında etkileşim yok demektir. "GRUP" satırındaki *p* değeri <0.05 (0.014) olduğuna göre dört farklı nedene bağlı effüzyonlarda ADA düzeyi ortalamalarının eşit olmadığı sonucuna varılabilir. Ayrıca "CINSIYET" satırındaki *p* değeri de <0.05 (0.002) olduğu için erkekler ve kadınlar arasında da ADA düzeyi ortalamalarının eşit olmadığı sonucuna varılabilir.

Tests of Between-Subjects Effects						
		Dependent Variable: ADA				
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	19132.590	1	19132.590	13.177	.171
	Error	1451.974	1	1451.974		
GRUP	Hypothesis	1796.317	3	598.772	23.890	.014
	Error	75.192	3	25.064		
CINSIYET	Hypothesis	1451.974	1	1451.974	53.642	.002
	Error	104.664	3	27.068		
GRUP * CINSIYET	Hypothesis	75.192	3	25.064	.383	.766
	Error	2095.116	32	65.472		

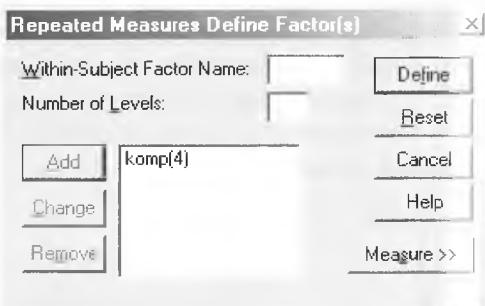
Şekil 12 Bağımlı değişken olarak ADA ve faktör değişkenleri olarak "GRUP" ve "CINSIYET"in alındığı iki yönlü varyans analizinin tablosu.

Univariate diyalog kutusundan yapılabilecek çok yararlı bir diğer analiz de kovaryans analizidir. Bununla ilgili ayrıntıları çok değişkenli istatistik analiz yöntemleri bölümünde (Bölüm 16) bulabilirsiniz.

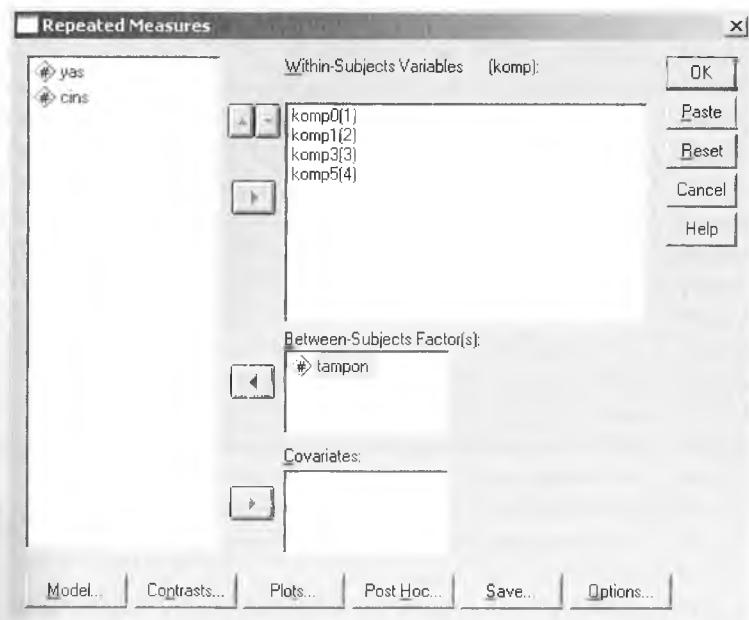
Tekrarlı ölçümler varyans analizi

Analyze >> General Linear Model >> Repeated Measures

Yukarıdaki seçenekle **Repeated Measures Define Factors** diyalog kutusu açılır (Şekil 13). *Within-Subject Factor Name* kutusuna, birden çok kez ölçülmüş olan bağımlı değişkenin (yani grupta değişkenin) adı ve *Number of Levels* kutusuna da ölçüm sayısı yazılmalıdır. Örnek 51'deki çalışma için *Within-Subject Factor Name* olarak komp ve *Number of Levels*'a da "4" yazın ve *Add* butonunu klikleyerek listeye alın. Daha sonra *Define* butonuna basıldığında **Repeated Measures** diyalog kutusu açılır (Şekil 14).



Şekil 13
SPSS'te "Repeated Measures Define Factors" diyalog kutusu.



Şekil 14
SPSS'te
“Repeated
Measures”
diyalog
kutusu.

Burada farklı zamanda ölçülmüş olan kompleman değerlerinin kaydedildiği komp0, komp1, komp3 ve komp5 adlı değişkenleri *Within-Subject Variables* listesine seçin. İki farklı tampon grubunun kodlandığı tampon değişkenini (yani gruplararası faktörü) ise *Between-Subjects Factor(s)* listesine seçin. Böylece iki farklı tampon kullanımının diyaliz seansı boyunca kompleman düzeyinin seyrine olan etkisinin inceleneceği tekrarlı ölçümler varyans analizi yapılabilecektir.

Analize başlamadan önce *Contrasts* altmenüsüne girip *Change Contrast*'tan *Simple*'ı (basit kontrast) ve *Reference Category* olarak da *First*'ü seçin. *Change* butonuna basarak kompleman değişkeninin kontrast tipini değiştirin. Böylece diyaliz sırasında kompleman değerleri 0. saat kompleman değerleri ile karşılaştırılacaktır. Eğer ardışık ölçütler arasındaki farkları (0. saat ile 1. saat, 1. saat ile 3. saat ve 3. saat ile 5. saat) incelemek istiyorsanız, kontrast tipi olarak *Repeated*'ı (ardışık kontrast) seçin.

Tanımlayıcı istatistikleri almak için *Options* altmenüsüne girip *Descriptive statistics*'i ve grafik çizdirmek için *Plots* altmenüsüne girip *Horizontal Axis*'e komp ve *Separate Lines*'a tampon'u seçin ve *Add* butonunu klikleyerek listeye alın.

Şekil 15'de kompleman kontrast tipinin referans ölçümün 0. saat değeri ol-

duğu basit kontrast olarak seçildiği analiz sonuçları görülmektedir. Şekil 16'da ise yukarıda tanımlanan grafik görülmektedir.

Şekil 15'de en üstteki tanımlayıcı istatistikler tablosunda kompleman düzeyi ortalamasının 0'inci saatte benzer olduğu, ancak daha sonraki ölçümlerde asetat grubunda yükselirken bikarbonat grubunda değişmediği görülmektedir. *Multivariate Tests* tablosunda KOMP yazan bölümde 4 farklı zamanda ölçülen kompleman düzeyleri arasında fark olup olmadığı, KOMP*TAMPON yazan bölümde ise tampon tipi ile komplemanın zamanla seyrini gösteren eğrilerin paralel olup olmadığı 4 farklı yöntemle test edilmiştir. *Sig* sütunundaki değerler, *p* değerleridir. Her ikisi için de *p* değeri <0.05 (0.000) olduğu için hem komplemanın 4 farklı zamanda ölçülen değerleri arasında fark olduğu, hem de kompleman eğrilerinin paralel olmadığı anlaşılmaktadır.

Mauchly's Test of Sphericity (Mauchly küresellik testi) tablosunda tekrarlı ölçümler varyans analizinin varsayımlarından biri olan bağımlı değişkenin kovaryans matriksinin varyansının sabit olup olmadığı test edilmiştir.

P değerinin >0.05 (0.054) olması küresellik varsayıminın sağlanmadığını göstermektedir. Bu varsayımin sağlanmaması tekrarlı ölçümler varyans analizinin geçersiz olduğu anlamına gelmez.

Küresellik varsayıminin sağlanamadığı durumlarda hesaplamalarda kullanılan serbestlik derecesi revize edilerek, geçerli sonuçlar elde edilir. *Test of Within Subjects Effects* tablosunda yer alan Greenhouse-Geisser, Huynh-Feldt ve Lower-bound testleri ile bu revizyon sağlanabilir. *Tests of Within-Subjects Effects* (Grup içi farklar) tablosunda *p* değerleri <0.05 (0.000) olduğu için hem komplemanın 4 farklı zamanda ölçülen değerleri arasında fark olduğu, hem de kompleman eğrilerinin paralel olmadığı anlaşılmaktadır.

Tests of Within-Subjects Contrasts (Grup içi kontrastlar) tablosunda 1, 3 ve 5'inci saat kompleman ortalamalarının 0'inci saat kompleman ortalamasından farklı olduğu, yine kompleman eğrilerinin 1, 3 ve 5'inci saatlerde paralel olmadığı anlaşılmaktadır. *Tests of Between-Subjects Effects* (Gruplar arası farklar) tablosunda iki tampon grubu arasında fark olduğu anlaşılmaktadır (*p*=0.000).

Descriptive Statistics				
	TAMPON	Mean	Std. Deviation	N
KOMP1	1 Asetat	63.95	8.50	20
	2 Bikarbonat	61.30	11.88	20
	Total	62.63	10.28	40
KOMP2	1 Asetat	82.50	16.61	20
	2 Bikarbonat	64.15	8.80	20
	Total	73.33	16.07	40
KOMP3	1 Asetat	119.45	22.21	20
	2 Bikarbonat	69.15	12.84	20
	Total	94.30	31.14	40
KOMP4	1 Asetat	106.70	16.81	20
	2 Bikarbonat	69.55	15.44	20
	Total	88.13	24.65	40

Şekil 15

Grupçi değişken olarak kompleman düzeyleri (basit kontrast) ve gruplararası değişken olarak tampon tipinin aldığı tekrarlı ölçümler varyans analizinin sonuçları. (sayfa 172'deki tablolara da bakınız).

Multivariate Tests						
Effect		Value	F	Hypothesis df	Error df	Sig.
KOMP	Pillai's Trace	.863	75.568	3.000	36.000	.000
	Wilks' Lambda	.137	75.568	3.000	36.000	.000
	Hotelling's Trace	6.297	75.568	3.000	36.000	.000
	Roy's Largest Root	6.297	75.568	3.000	36.000	.000
KOMP * TAMPON	Pillai's Trace	.769	39.907	3.000	36.000	.000
	Wilks' Lambda	.231	39.907	3.000	36.000	.000
	Hotelling's Trace	3.326	39.907	3.000	36.000	.000
	Roy's Largest Root	3.326	39.907	3.000	36.000	.000

Mauchly's Test of Sphericity							
Measure: MEASURE_1							
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
KOMP	.744	10.876	5	.054	.868	.962	.333

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
KOMP	Sphericity Assumed	24651.669	3	8217.223	88.066	.000
	Greenhouse-Geisser	24651.669	2.604	9467.012	88.066	.000
	Huynh-Feldt	24651.669	2.886	8541.378	88.066	.000
	Lower-bound	24651.669	1.000	24651.669	88.066	.000
KOMP * TAMPON	Sphericity Assumed	13136.069	3	4378.690	46.928	.000
	Greenhouse-Geisser	13136.069	2.604	5044.661	46.928	.000
	Huynh-Feldt	13136.069	2.886	4551.421	46.928	.000
	Lower-bound	13136.069	1.000	13136.069	46.928	.000
Error(KOMP)	Sphericity Assumed	10637.012	114	93.307		
	Greenhouse-Geisser	10637.012	98.950	107.499		
	Huynh-Feldt	10637.012	109.674	96.988		
	Lower-bound	10637.012	38.000	279.921		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

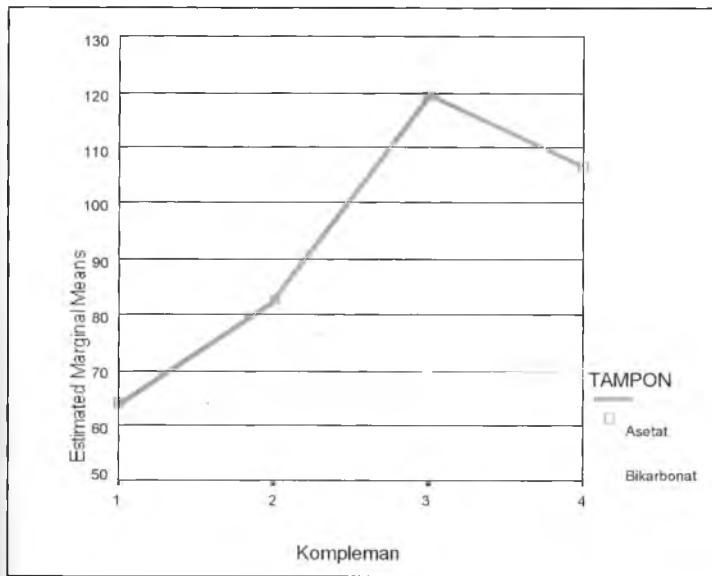
Source	KOMP	Type III Sum of Squares	df	Mean Square	F	Sig.
KOMP	Level 2 vs. Level 1	4579.600	1	4579.600	48.428	.000
	Level 3 vs. Level 1	40132.225	1	40132.225	195.177	.000
	Level 4 vs. Level 1	26010.000	1	26010.000	144.934	.000
KOMP * TAMPON	Level 2 vs. Level 1	2464.900	1	2464.900	26.065	.000
	Level 3 vs. Level 1	22705.225	1	22705.225	110.423	.000
	Level 4 vs. Level 1	11902.500	1	11902.500	66.324	.000
Error(KOMP)	Level 2 vs. Level 1	3593.500	38	94.566		
	Level 3 vs. Level 1	7813.550	38	205.620		
	Level 4 vs. Level 1	6819.500	38	179.461		

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	253406.6	1	253406.6	1709.446	.000
TAMPON	7350.877	1	7350.877	49.588	.000
Error	5633.084	38	148.239		



Şekil 16
Asetat ve bikarbonat tamponla diyaliz yapılan hastalarda 4 farklı zamanda ölçülen kompleman düzeyi ortalamalarının gösterildiği çizgi grafik.

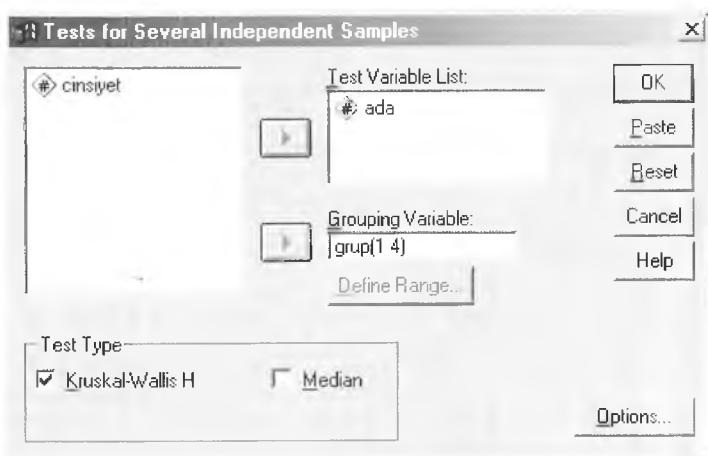
Kruskal-Wallis tek yönlü varyans analizi

Analyze >> Nonparametric Tests >> K Independent Samples

Yukarıdaki seçimle **Tests for Several Independent Samples** diyalog kutusu açılır (Şekil 17). Bağımlı değişken ya da değişkenleri *Test Variable List'e*, grupları belirleyen değişkeni yani faktörü ise *Grouping Variable* kutusuna seçin ve *Define Range* seçeneği ile minimum ve maksimum kategori kodlarını belirleyin.

Şekil 18'de effüzyon gruplarının ADA ortancalarının karşılaştırıldığı Kruskal-Wallis analizinin sonucu gösterilmiştir. *Sig* olarak gösterilen p değeri <0.05 (0.004) olduğuna göre grupların ADA ortancalarının farklı olduğu sonucuna varılmıştır. Bu durumda, ikişerli karşılaştırmalar Mann-Whitney U testi ile yapılabilir. Ancak istatistiksel önemlilik sınırı olarak p değerinin 0.05'den küçük olması değil, $0.05/6 = 0.008$ 'den küçük olmasına dikkat etmelidir.

Şekil 17
SPSS'te
"Nonparametric
tests"
seçeneginden
"Tests for
Several
Independent
Samples" diyalog
kutusu.



Şekil 18
Dört farklı nedene bağlı plevral effüzyonda ADA
düzeyi ortancalarının karşılaştırıldığı
Kruskal-Wallis testinin sonuçları.

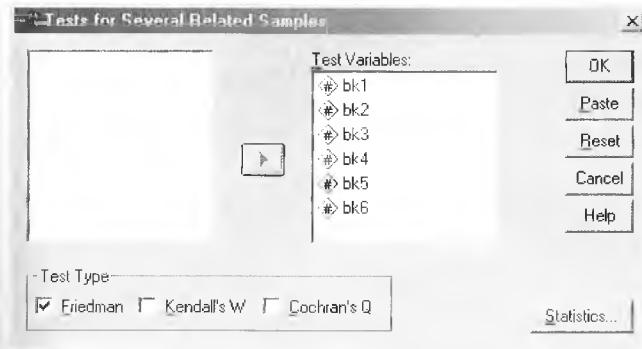
Test Statistics ^{a,b}	
	ADA
Chi-Square	13.523
df	3
Asymp. Sig.	.004

^a Kruskal Wallis Test
^b Grouping Variable: GRUP

Friedman iki yönlü varyans analizi

Analyze >> Nonparametric Tests >> K Related Samples

Yukarıdaki seçimle Tests for Several Related Samples diyalog kutusu açılır (Şekil 19). Bağımlı değişkenleri *Test Variables* kutusuna seçin ve OK'leyin.



Şekil 19
SPSS'te "Nonparametric Tests" seçenekinden "Tests for Several Related Samples" diyalog kutusu.

Şekil 20'de 16 hastada altı farklı zamanda ölçülen beyaz küre (BK1, BK2, ..., BK6) değerlerinin ortancalarının karşılaştırıldığı Friedman analizinin sonucu gösterilmiştir. *Sig* olarak gösterilen *p* değeri <0.05 (0.000) olduğuna göre beyaz küre ortancaları eşit değildir. İkişerli karşılaştırmalar Wilcoxon testi ile yapılabilir. Ancak istatistiksel önemlilik sınırı, yapılacak ikişerli karşılaştırma sayısı ile ilişkili olarak değiştirilmelidir.

Örneğin başlangıçtan sonraki ortancalar (BK2, ..., BK6), başlangıçtaki ortanca (BK1) ile karşılaştırılacaksa, toplam ikişerli karşılaştırma sayısı beş olacağı için, istatistiksel önemlilik sınırı, *p* değerinin 0.05'den küçük olması değil, $0.05/5 = 0.01$ 'den küçük olması olarak tanımlanmalıdır.

Test Statistics ^a	
N	16
Chi-Square	74.354
df	5
Asymp. Sig.	.000

a. Friedman Test

Şekil 20

Altı farklı zamanda ölçülen beyaz küre değerlerinin karşılaştırıldığı Friedman testinin sonuçları.

ORANLARIN KARŞILAŞTIRILMASI

bölüm 12

Tek grup oranı

Güven sınırları

Hipotez testi

Bağımsız oranların karşılaştırılması

Güven sınırları

Z testi

Ki-kare testi

Çapraz tablolar

Ki-kare testinin varsayımları

Fisher kesin testi

2x2'lik tablolar için ki-kare düzeltmesi

İkiden fazla bağımsız grup oranının karşılaştırılması

2x2'den büyük çapraz tablolar için ki-kare testi

P değeri <0.05 ise ne yapmalı?

Beklenen sayılar küçükse ne yapmalı?

Relatif risk (risk oranı)

Odds oranı

Risk oranı mı, odds oranı mı?

Risk oranı mı, ki-kare mi?

Bağımlı oranların karşılaştırılması

McNemar testi

Cochran's Q testi

Marjinal homojenlik testi

Kappa istatistiği

SPSS'te bağımsız grup oranları nasıl karşılaştırılır?

Crosstabs menüsü

Cells altmenüsü

Statistics altmenüsü

Ki-kare (χ^2) (chi-square)

Risk

SPSS'te bağımlı grup oranları nasıl karşılaştırılır?

McNemar testi

Kappa istatistiği

Cochran Q testi

Marjinal homojenlik testi

TEK GRUP ORANI

Sayısal olarak ölçülen değişkenlerin ortalamalarının karşılaştırılmasında t testleri ve varyans analizinin kullanıldığı ve bu testlerde hesaplanan t değerleri ile tablodaki t değerlerinin karşılaştırıldığı daha önceki böltümlerde anlatılmıştı. Eğer değişkenlerin ölçüm skalası nominalse ya da ordinalse ortalamalar değil, oranlar karşılaştırılır. Tek grup, bağımsız iki grup ya da bağımsız iki den fazla gruba ait oranlar karşılaştırılabileceği gibi, önce-sona düzende yapılmış çalışmalarda bağlılı iki grup oranları da test edilebilir.

ÖRNEK 52

Tıpta tek grup içeren klinik çalışmalar pek sık kullanılmaz. Beta-bloker kullanımının miyokard infarktüsü sonrası ölümleri azalttığı varsayımlını araştırmak için bir çalışmada tek grup kullanıldığını düşünelim. Her birinde 50 hasta bulunan iki gruptan birine beta-bloker, diğerine placebo verilerek, iki gruptaki ölüm oranları karşılaştırılıyor. Eğer kontrol grubu kullanılmazsa, yani çalışmaya alınan tüm hastalara beta-bloker verilerek bir çalışma yapılırsa, çalışma grubunun ölüm oranının, literatürden bilinen ya da aynı hastanede daha önceki bir çalışmadan elde edilen başka bir ilaca ait ölüm oranı ile karşılaştırılması yoluna başvurulabilir. Bu örnekte ilaca yanıt, ölüm olması ya da olmaması şeklinde; yani nominal değer almaktadır.

Güven sınırları

Tek çalışma grubundaki bir nominal değişkenin görülmesi oranı p ve toplam denek sayısı n ise, np ve $n(1-p)$ 5'ten büyükse, değişkenin dağılımı normaldir. Örnek 52'de 50 hastada ölüm oranı %12 ise $np = 50 \times 0.12 = 6$, ve $n(1-p) = 50 \times 0.88 = 44$ olduğu için, yani her ikisi de 5'ten büyük olduğu için grup oranının güven sınırları hesaplanabilir. Dikkat edilirse ölen hasta sayısı 6, yaşayan hasta sayısı 44'tür. Yani aslında %12 ölüm oranı bu rakamlardan hesaplanmıştır. Eğer ölen hasta sayısı ya da yaşayan hasta 5'ten azsa, tek grup bun güven sınırlarının doğru olarak hesaplanması mümkün olmaz.

Tek grup oranının %95 güven sınırları aşağıdaki formülle hesaplanabilir:

$$\%95 \text{ güven sınırları} = p \pm 1.96 \sqrt{\frac{p(1-p)}{n}}$$

Örnek 52'de $n = 50$ ve $p = 0.12$ 'dir. %95 güven aralığı hesaplandığında,

$$0.12 \pm (1.96) \sqrt{\frac{(0.12)(0.88)}{50}} = 0.12 \pm 0.09 \text{dur, yani ölüm oranı \%95 olasılıkla}$$

%3 ile %21 arasındadır.

Eğer daha önce kullanılan tedavi yöntemi ile gözlenen ölüm oranı %22 ya da daha fazlaysa, beta-blokerin ölüm oranını önemli derecede azalttığı sonucuna varılabilir.

Hipotez testi

Örnek 52'deki çalışmada %12 olarak bulunan ölüm oranının kalsiyum antagonistlerinin kullanıldığı başka bir çalışmada elde edilen %25'ten daha düşük olup olmadığı da %95 güven sınırlarının belirlenmesine benzer olarak test edilebilir. Bu durumda H_0 hipotezi "beta-bloker çalışmasındaki ölüm oranı ≥ 0.25 ", H_1 hipotezi "beta-bloker çalışmasındaki ölüm oranı < 0.25 " olarak kurulur. Dikkat edilirse, H_0 hipotezinde " $=0.25$ " yerine " ≥ 0.25 " kullanılmıştır. Bu nedenle iki-yönlü değil, tek yönlü test uygulanacaktır.

Hipotez testinin uygulanabilmesi için $n\pi$ ve $n(1 - \pi)$ 'nin 5'ten büyük olması gereklidir. Burada π yerine örneklem oranı olan 0.12 değil, evren oranı olan 0.25 konulmalıdır. Hesaplandığında bu iki değer 12.5 ve 37.5'tür. Her ikisi de 5'ten büyük olduğu için, hipotez testi uygulanabilir.

Hipotez testinde aşağıdaki formülle Z değeri hesaplanır.

$$Z = \frac{p - \pi}{\sqrt{\pi(1-\pi)/n}}$$

Bu formüle göre hesaplandığında, $Z = \frac{0.12 - 0.25}{\sqrt{0.25(1-0.25)/50}} = -2.12$ bulunur. α

yanılma düzeyi 0.05'e karşılık gelen Z değeri tek yönlü test için 1.645'tir. Hesaplanan z değerinin mutlak değeri olan 2.12, tablodaki Z değerinden büyük olduğu için H_1 hipotezi kabul edilir. Yani %95 olasılıkla beta-bloker kullanımı sonrası gözlenen %12 ölüm oranı, %25'ten daha düşüktür.

BAĞIMSIZ ORANLARIN KARŞILAŞTIRILMASI

Örnek 52'de çalışmaya alınan tüm hastalara beta-bloker verip de, sonuçları daha önce yapılan bir çalışmaya karşılaştırırmak, farklı çalışmalar için hastalık kriterleri farklı olabileceği için, hatalı yorumlara neden olabilir. Bu ne-

denle, çalışmaya bir kontrol grubu eklenmesi yararlı olur. Hastalığın özelliğine göre kontrol grubuna placebo verilebilir, ya da etik nedenlerle tedavisiz bırakılamayacak bir hastalıkla, başka bir tedavi verilebilir.

ÖRNEK 53

Miyokard infarktüsü sonrası tedavi verilmemesi etik olmayacağı için 50 hastaya beta-bloker ve 50 hastaya etkisi bilinen başka bir ilaç olarak kalsiyum antagonisti verildiğini varsayılm. Bu çalışma düzeni bilindiği gibi, bağımsız eşzamanlı kontrollü klinik çalışma düzenine bir örnektir.Tİpta en sık kullanılan çalışma düzenidir.

Bağımsız iki grup oranının karşılaştırılmasında güven sınırları, z-testi ve ki-kare testi olmak üzere üç farklı yöntem seçilebilir. Üçünden de elde edilecek sonuç aynıdır.

Güven sınırları

Elli hastaya beta-bloker, 50 hastaya da kalsiyum antagonisti verildiğini ve her iki grupta ölüm oranlarının sırasıyla %12 ve %24 olarak bulunduğu varsayılm. İki ilaç arasında ölüm oranları açısından fark olup olmadığını test etmek için iki oran arasındaki farkın %95 güven sınırları hesaplanabilir. Eğer hesaplanan alt güven sınırı %0'ı içermiyorsa, iki oran arasındaki farkın önemli olduğu söylenebilir. Güven sınırlarının hesaplandığı formül karışık olduğu için verilmemiştir.

Bağımsız iki grup oranı arasındaki farkın güven sınırlarının hesaplanabilmesi için her iki grup için de np ve $n(1-p)$ 'nin 5'ten büyük olması gereklidir. Bu değerler beta-bloker grubu için 6 ve 44, kalsiyum antagonisti için 12 ve 38 olduğunu göre güven sınırları hesaplanabilir.

Z testi

Bağımsız iki grup oranının karşılaştırılmasında güven sınırları yerine Z testi de kullanılabilir. Bu testin uygulanabilmesi için de np ve $n(1-p)$ 'nin 5'ten büyük olması koşulu vardır.

Bu testte grup oranları ve denek sayılarından hesaplanan Z değerinin mutlak değeri, tablodaki Z değerinden büyükse, grup oranlarının birbirinden farklı olduğu sonucuna varılabilir. Tablodan Z değerine bakılırken H_1 hipotezinin yapısına dikkat edilmelidir. Eğer H_1 hipotezi "beta-bloker grubunun ölüm oranı, kalsiyum antagonisti grubunun ölüm oranından düşüktür" ise

tek yönlü, "beta-bloker grubunun ölüm oranı, kalsiyum antagonistı grubunun ölüm oranından farklıdır" ise iki yönlü Z değeri alınmalıdır.

Ki-kare testi

Bağımsız grup oranlarının karşılaştırılmasında en sık kullanılan test ki-kare (χ^2) testidir. Z testinden farklı olarak ikiden çok grup oranları da karşılaştırılabilir. Aslında Z testinde iki grup oranının birbirinden farklı olup olmadığı sorusuna yanıt aranırken, ki-kare testinde bir nominal değişkenin (örneğin ilaç tipinin) diğer bir nominal değişkenle (örneğin прогнозla) ilişkili olup olmadığı araştırılır. Değişkenlerden biri ya da ikisinin ordinal olduğu durumlarda da ki-kare testi uygulanabilir. Ki-kare testinin ayrıntılarına girmeden önce nominal ve ordinal verilerin özetlendiği çapraz tablo adı verilen tablolara ilişkin bazı bilgilerin verilmesinde yarar var.

Çapraz tablolar

Ölçüm düzeyleri "Evet, Hayır", "Erkek, Kadın", "Karadeniz, İç Anadolu, Ege, ..." gibi nominal, ya da "Evre I, II, III, IV", "Hayır, Belki, Evet" gibi ordinal olan değişkenlerin ortalama, ortanca gibi tanımlayıcı istatistikleri yerine "erkeklerin oranı, İç Anadolu bölgesinden gelenlerin oranı" ya da daha basitçe "her iki cinsteki deneklerin sayısı, her bölgeden gelen kişilerin sayısı" gibi değerler verilebilir. Nominal ya da ordinal değerler alan değişkenlerin değerlerinin gösterildiği tablolara çapraz tablolar adı verilir. Çapraz tabloların yapılması oldukça basittir, ama çok yararlı ve kolay anlaşılır bir tablo türüdür.

ÖRNEK 54

Romatoid artritli (RA) ve normal kişilerde analjezik kullanımının araştırıldığı bir çalışmada eğitim düzeyinin de analjezik kullanma davranışını etkileyebileceği düşünülerek, grupların eğitim açısından farklı olup olmadığı inceleniyor. Denegin RA grubuna mı, normal gruba mı ait olduğu "RA ve kontrol" olarak, eğitim durumu ise "okumamış, ilk-orta-lise ve yüksek okul" olarak kodlanmış. Yani GRUP değişkeni iki düzeyli nominal, EĞİTİM değişkeni ise üç düzeyli ordinal değişkenlerdir. Çalışmanın sonuçları Tablo 1'deki çapraz tabloda özetlenmiştir.

Tablo 1'deki 3X2'lik çapraz tabloda görüldüğü gibi altı göz vardır. Satır değişkeni GRUP, sütün değişkeni EĞİTİM'dir. Gözlerdeki ilk rakamlar denek yüzdesini, ikinci rakamlar satır yüzdesini, üçüncü rakamlar sütün yüzdesini ve son rakamlar ise toplam yüzdeyi göstermektedir.

Tablo 1 Hastalık grubu ve eğitim durumu değişkenlerinin 3X2'lik çapraz tablosu.

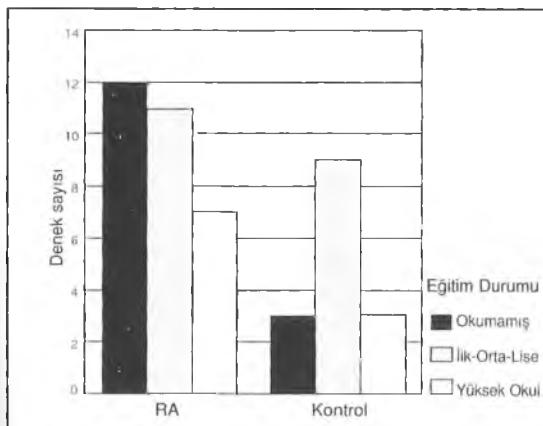
Sayı Satır yüzdesi Sütun yüzdesi Toplam yüzde	Eğitim Durumu			
	Okumamış	İlk-Orta-Lise	Yüksek Okul	Satır toplamı
RA	12	11	7	30
	40.0	36.7	23.3	66.7
	80.0	55.0	70.0	
	26.7	24.4	15.6	
Kontrol	3	9	3	15
	20.0	60.0	20.0	33.3
	20.0	45.0	30.0	
	6.7	20.0	6.7	
Sütun toplamı	15	20	10	45
	33.3	44.4	22.2	100.0

Satır yüzdesi, o gözdeki deneklerin satır toplamına oranını ifade etmektedir. Örneğin okumamış RA hastalarının RA hastaları toplamına oranı %40.0 olarak görülmektedir. Sütun yüzdesi ise o gözdeki deneklerin sütun toplamına oranını verir. Örneğin yüksek okul okumuş kişilerin %70.0'ı RA grubundadır.

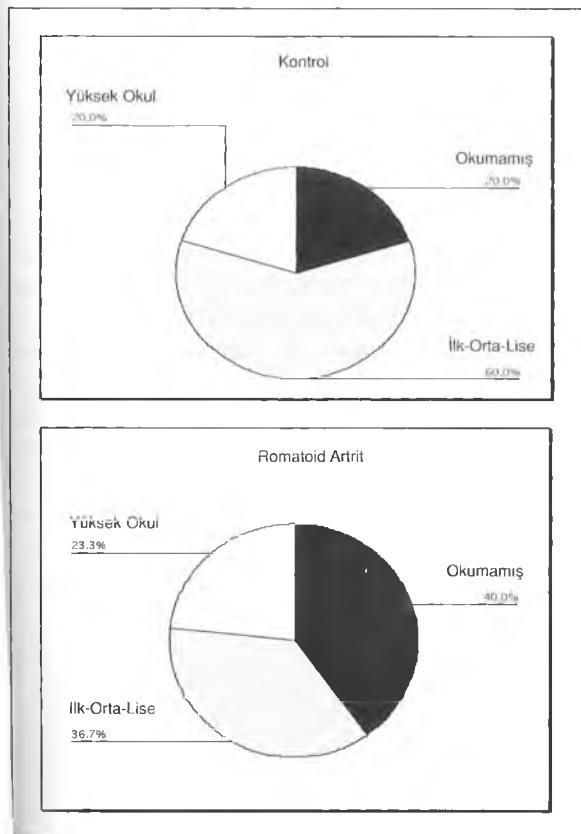
Toplam yüzde ise, o gözdeki deneklerin toplam deneklerin yüzde kaç olduğu anlamına gelir. Örneğin okumamış RA hastalarının, RA ve kontrol grubu karışık olmak üzere toplam gruba oranı %26.7'dir. Genellikle ya satır yüzdesi ya da sütun yüzdesi merak edilen orandır. Tablo 1'deki örnekte eğitim durumunun RA ile kontrol grubu arasında farklı olup olmadığı ile ilgilenilmesi için, grup değişkeni bağımsız, eğitim durumu değişkeni ise gruba bağlı değişken olarak tanımlanır.

Bağımsız değişken satır değişkeni ise, satır yüzdeleri, sütun değişkeni ise sütun yüzdeleri değerli, diğer yüzdeler ise genellikle yararsız, hatta yanlıştır. Örneğin bu çalışmadan RA hastalarının %40'ının okumamış olduğunu söylemek yanlış olmaz, ama yüksek okul mezunlarının %70'inin RA olduğunu yanlış olduğu çok açıklıktır.

Çapraz tablolar, nominal ve ordinal verilerin sunulması için çok yararlı yöntemlerdir. Aynı amaç için daha görsel amaçlı olarak çubuk grafik ve dilimli grafik çizilebilir (Şekil 1 ve 2).



Şekil 1
Tablo 1'deki
verilerin çubuk
grafikle gösterimi.



Şekil 2
Şekil 1'deki verilerin
dilimli grafikle gösterimi.

Ki-kare testine geçmeden önce çapraz tablolarla ilgili terimleri tanımlamakta yarar var. Örnek 53'deki çalışmanın sonuçları Tablo 2'deki şekilde bir çapraz tabloda gösterilebilir.

Tablo 2 Örnek 53'deki çalışmanın sonuçlarının gösterildiği çapraz tablo.

	Öldü	Yaşıyor	Toplam
Beta-bloker	6	44	50
Kalsiyum antagonistisi	12	38	50
Toplam	18	82	100

Bu tabloda en alta satır ve en sağdaki sütündaki rakamlara kenar sayıları ya da marjinal frekanslar adı verilir. Bu rakamlara bakıldığından çalışmaya 100 hasta alındığı, bunların 50'sine beta-bloker, 50'sine kalsiyum antagonistisi verildiği ve 100 hastadan 18'inin öldüğü anlaşılmaktadır. Eğer her iki ilaç arasında ölüm oranı açısından fark olmasaydı, beta-bloker grubunda 9 ve kalsiyum antagonistisi grubunda da 9 kişinin ölmüş olması beklenacaktı. Bu değerlere beklenen sayılar (*expected frequencies*) adı verilir. Bu grplardaki gerçek ölen kişi sayısı olan 6 ve 12 ise gözlenen sayılar (*observed frequencies*) olarak adlandırılır. Gözlenen ve beklenen sayılar ne kadar birbirinden farklısa, ölüm oranı ile kullanılan ilaç arasında o derece yakın ilişki var demektir.

Ki-kare testinde her gözde gözlenen ve beklenen sayıların farklarının karesi beklenen değere bölünür ve tüm gözlerden elde edilen rakamlar toplanarak, ki-kare değeri hesaplanır. 2×2 'lik çapraz tabloda 4 göz olduğu için ki-karenin hesaplanmasında 4 terim yer alır. Hesaplanan ki-kare değeri, tablodan alınan ki-kare değerinden büyükse, ölüm oranları ile kullanılan ilaç arasında o derece yakın ilişki var demektir. Tabloda serbestlik derecesi olarak (satır sayısı - 1) (sütun sayısı - 1) dikkate alınır. Bu nedenle 2×2 'lik çapraz tablo için serbestlik derecesi $(2 - 1)(2 - 1) = 1$ 'dir.

$$\chi^2 = \sum \frac{(G - E)^2}{E}$$

$$\text{Serbestlik derecesi} = (r - 1)(c - 1)$$

Bu formülde G gözlenen sayıyı, B beklenen sayıyı, r satır sayısını ve c sütun sayısını simgelemektedir.

Kolay hesaplanması ve anlaşılabilirliği nedeniyle ki-kare testi tip literatüründe en sık kullanılan testlerden biri olmuştur. Bu nedenle bazen de daha uygun istatistik analiz yöntemleri yerine ki-kare testi kullanılmaktadır. Ki-kare testi, iki nominal değişken arasındaki ilişkiyi inceler. Örnek 53'deki çalışmada ki-kare testi ile aralarındaki ilişkinin araştırıldığı iki nominal değişkenden biri verilen ilaç (beta-bloker ya da kalsiyum antagonisti) diğeri ise prognozdur (kişinin ölümü ya da yaşamaya devam etmesi). Sayısal değişkenlerin kategorize edilerek nominal ya da ordinal değişkenlere dönüştürülmesi ile veri kaybına neden olunabileceği daha önceki bölümlerde anlatılmıştır. Aşağıdaki çalışmanın ki-kare testi ile analizi bu tip hataya bir örnektir.

ÖRNEK 55

Örnek 53'deki çalışmada iki grup arasında ölüm oranı açısından değil de, 24 saatte oluşan ventriküler ekstrasistol (VES) sayısı açısından fark olup olmadığıının araştırılmak istendiğini varsayıyalım. Bu durumda değişkenlerden biri nominal (verilen ilaç), diğeri ise sayısal (VES sayısı) olacaktır.

Örnek 55'deki çalışmanın analizinde iki gruptaki VES sayısı ortalamalarının t testi ile karşılaştırılması doğru olan yöntemdir. Ancak bilindiği gibi sayısal değişkenler kategorize edilerek, ordinal ya da ikili değişkenlere dönüştürülebilir. Bu örnekte VES sayısının "24 saatte 50'den az" ve "50'ye eşit ya da daha fazla" olarak ikili değişkene dönüştürülmesi ile aşağıdaki çapraz tablo elde edilebilir (Tablo 3). Bu çapraz tabloya ki-kare testi uygulandığında verilen ilaçla VES sayısı arasında ilişki olmadığı sonucu çıkmaktadır.

Tablo 3 Örnek 55'deki çalışmanın sonuçlarının verildiği çapraz tablo.

	VES sayısı		Toplam
	<50/gün	≥50/gün	
Beta-bloker	15	35	50
Kalsiyum antagonistisi	11	39	50
Toplam	26	74	100

Sayısal değişkenlerin kategorize edilerek ikili ya da ordinal değişkenlere dönüştürüldü, daha sonra ki-kare testi yapılması ile önemli ölçüde veri kaybı olmuştur. Örneğin VES sayısı günde 50'den fazla olan ve beta-bloker grubunda bulunan 35 hastada VES sayısının genellikle 51-100/gün arasında, kalsiyum antagonist grubu bulunan 39 hastada ise genellikle 500-1000/gün arasında olduğunu varsayılmıştır. Kalsiyum antagonist grubunda günlük VES sayısı daha fazla olduğu halde, VES sayısının kategorize edilmesi ile bu sonuca ulaşılacak veriler kullanılmamış olduğu için her iki ilaç grubunda da VES sayısı açısından fark olmadığı sonucu çıkmıştır. Bu durum için en uygun istatistik yöntem, iki ilaç grubunun VES sayıları ortalamaları arasında fark olup olmadığını t testi ile araştırılmasıdır. Bu yöntemle beta-bloker grubunda VES sayısının daha az olduğu gösterilebilir.

Sayısal değişkenleri kategorize etmekle veri kaybı olduğu kesin olduğu halde, bazen klinik pratikteki bilgiler nedeniyle yukarıdaki yöntem uygulanabilir. Örneğin VES sayısının günde 50'den az olması ile ani ölüm riskinin azaldığını ve 50'den fazla olduğu takdirde ister 60, ister 360 olsun fark olmaksızın ani ölüm riskinin aynı olduğunu varsayılmıştır. Bu durumda iki ilaç grubundaki VES sayıları ortalamaları arasında fark olduğu halde, VES sayısının günde 50'nin altında olduğu hastaların oranlarının farksız olduğunun bilinmesi yanlış olacaktır.

Sonuç olarak, eğer sayısal bir değişkenin tıbbi nedenlerle belirlenmiş bir sınır değeri (*cutoff*) varsa, verilerin kategorize edilmesi yararlı olabilir. Bunun dışında sayısal değişkenleri kategorize etmek veri kaybına yol açar.

Ki-kare testinin varsayımları

Ki-kare testinin en önemli varsayıımı, oranların karşılaştırıldığı grupların bağımsız olmasıdır. Örnek 55'deki çalışmaya alınan hastalardan bazıları farklı zamanlarda beta-bloker ve kalsiyum antagonisti kullanmışsa ve ilacın etkilerinin değerlendirilmesinde bu hastaların her iki ilacı kullandığı dönemdeki veriler farklı gruplara ait veriler olarak ele alınırsa, ki-kare testinin "grupların bağımsızlığı" varsayıımı gerçekleşmemiş olur. Bu nedenle bir grupta yer alan bir denegin diğer grupta da verisinin bulunmaması gereklidir.

Z testinin uygulanabilmesi için n ve $n(1-p)$ 'nin 5'ten büyük olması gereği gibi ki-kare testinin uygulanabilmesi için de beklenen sayıların çok küçük olmaması gereklidir. Formüle dikkat edilirse, herhangi bir gözdeki beklenen sayı ne kadar küçük olursa, ki-kare değeri o kadar büyük olacak ve iki değişken arasında belki de gerçekte ilişki olmadığı halde ilişki varmış gibi sonuç çıksa caktır. Gözlerdeki beklenen sayılardan herhangi birinin 2'den küçük ya da gözlerin yarısından çoğundaki beklenen sayıların 5'ten küçük olması durum-

munda ki-kare testinin yapılması doğru değildir. Bazı istatistik yazarları bu konuda daha konservatif davranışarak, herhangi bir gözdeki beklenen değerin 5'ten küçük olması durumunda ki-kare testinin yapılmasının hatalı olacağını öne sürmektedir.

Örnek 53'te beta-bloker grubunda 6 kişi değil de, 1 kişinin öldüğünü varsayıyalım. Bu durumda Tablo 4'deki çapraz tablo elde edilecektir.

Tablo 4 Örnek 53'teki çalışmada farklı ölüm oranları ile elde edilen çapraz tablo.

	Öldü	Yaşıyor	Toplam
Beta-bloker	1	49	50
Kalsiyum antagonisti	12	38	50
Toplam	13	87	100

Beklenen sayılar, "Öldü" sütunu için 6.5 ve 6.5, "Yaşıyor" sütunu için 43.5 ve 43.5 olacaktır. Dört gözdeki beklenen sayılar da 5'ten büyük olduğu için ki-kare testi yapılabilir.

Gözlenen sayılar 5'ten küçük bile olsa, beklenen sayılar yeterince büyükse ki-kare testi yapılabilir.

Fisher kesin testi

Beklenen sayılar küçükse, ki-kare testi yerine Fisher kesin testi uygulanabilir. Fisher testinde doğrudan p değeri hesaplanır. Elle hesaplanması oldukça zordur. Ki-kare testi ile 2X2'den daha büyük çapraz tablolar için de değişkenler arası ilişki analizi yapılabılırken, Fisher testi yalnız 2X2'lik çapraz tablolarda kullanılabilir.

Tablo 1'den 30 RA hastasının 12'sinin (%40), 15 normal deneğin ise 3'ünün (%20) okumamış olduğu anlaşılmaktadır. RA ve kontrol grubu birleşik olarak bakıldığından ise sayısı 45 olan tüm deneklerin 15'inin (%33.3) okumamış olduğu görülmektedir. Eğer iki grup tamamen birbirine benzeseydi, acaba 30 RA hastasından kaçının okumamış olması beklenirdi? Bu soruya yanıt bulmak kolay. Toplam 45 kişiden 15'i okumamış olduğuna göre, 30 RA hastasından 10'unun okumamış olması beklenirdi. O halde RA-okumamış gözündeki gözlenen değer 12 olduğu halde, beklenen değer 10'dur. Benzer şekilde tüm gözler için beklenen değerler bulunabilir.

Beklenen değerlerin, gözlenen değerler gibi tam sayı olması gerekmekz. Gözlerdeki beklenen değerler, gözlenen değerlerden ne kadar farklıysa, grup-

lar arasında fark olma olasılığı da o kadar fazla olacaktır. Ki-kare testi bu temele dayanır. Ki-kare testinde hesaplanan χ^2 değeri ve göz sayıları dikkate alınarak hesaplanan p değeri <0.05 ise, "gözlerden en az birisinde gözlenen değerle beklenen değer arasında önemli fark var" demektir.

Denek sayıları azsa, ki-kare testinin güvenilirliği azalır. Bu nedenle herhangi bir gözdeki beklenen değer 5'ten küçükse, ki-kare testi yapılmaması önerilir. Bu durumda 2X2'lik çapraz tablolar için Fisher testi uygulanmalıdır. Satır ya da sütun sayıları ikiden fazlaysa, Fisher testi yapılamaz. Bu durumda ikiden fazla olan satır ya da sütun düzeylerinin birleştirilmesi yoluna gidilmelidir. Örneğin Tablo 1'deki gibi 3X2'lik bir tabloda beklenen değerlerin küçük olması nedeniyle ki-kare testi yapılamazsa, eğitim durumu değişkeninin düzey sayısı üçten ikiye indirilmelidir. Örneğin denekler "okur-yazar değil ve okur yazar" olarak belirlenebilir.

Böylece oluşturulan 2X2'lik çapraz tabloda ya ki-kare testinin yapılabilmesi koşulu olan beklenen değerlerin 5'ten büyük olması koşulu sağlanmış olur -ki bu durumda ki-kare testi yapılabılır-, ya da eğer hala beklenen değerlerden en az biri 5'ten küçükse, tablo 4 gözlü olduğu için Fisher testi yapılabilir. Fisher testinde beklenen değer sayısının büyülüüğünün önemi yoktur. Fisher testinde t, F, χ^2 gibi bir test istatistiği yoktur. Doğrudan p değerini verir.

ÖRNEK 56

**İki farklı antihipertansifin kan basıncını düşürücü etkisinin
araştırıldığı bir çalışmada Adipin ilacının verildiği 40 hastanın 38'inde,
Bedipin ilacının verildiği 60 hastanın 22'sinde ilaç etkili bulunmuştur. Gözler
den birindeki denek sayısı 2. Ki-kare testi yapılabilir mi?**

Bu soruyu yanıtlamak için her gözdeki beklenen değeri bulmak gereklidir (Tablo 5). Gözlenen değere bakarak beklenen değeri kestirmek olanaksızdır. Nitekim Adipine yanıt vermeyen hasta sayısı 2 olmasına karşın, ilaçlar arasında fark olmasaydı, 16 hastanın yanıt vermeyeceği beklenirdi. Beklenen değer 5'ten büyük olduğu için, ki-kare testi yapılabilir.

Tablo 5 Adipin ve Bedipin'in kan basıncı düşürücü etkisinin çapraz tablosu.

Gözlenen değer (Beklenen değer)	Kan basıncı normale düştü mü?		Toplam
	Evet	Hayır	
Adipin	38 (24)	2 (16)	40
Bedipin	22 (36)	38 (24)	60
Toplam	60	40	100

ÖRNEK 57

Oral ve parenteral demir preparatlarının allerjik yan etkileri olup olmadığını araştırıldığı bir çalışmada oral demir preparatının verildiği 80 hastanın 6'sında, parenteral demir preparatının verildiği 20 hastanın da 6'sında yan etki ortaya çıkmıştır. İki preparat arasında allerjik yan etki açısından fark olup olmadığı nasıl anlaşılabılır?

Bu soruyu yanıtlamak için de her gözdeki beklenen değeri bulmak gerekir (Tablo 6). Gözlenen değere bakarak beklenen değeri kestirmek olanaksızdır. Niçin ki her iki grupta da yan etki gözlenen denek sayısı 6 olduğu halde, oral-parenteral demir preparatları arasında fark olmasaydı, parenteral tedavi verilen grupta, 2.4 hastada yan etki olması beklenebilirdi. Beklenen değer 5'ten küçük olduğu için, ki-kare testi yapılması sakincalı olabilir. Fisher testi yapılmalıdır.

Tablo 6 Demir preparatının tipi ve allerjik yan etki olup olmadığına ilişkin çapraz tablo.

Gözlenen değer (Beklenen değer)	Allerjik yan etki		Toplam
	Yok	Var	
Oral	74 (70.4)	6 (9.6)	80
Parenteral	14 (17.6)	6 (2.4)	20
Toplam	88	12	100

Ki-kare testi satır ve sütun değişkenlerinin birbirinden bağımsız olup olmadığını test eder. Bulunan p değeri >0.05 ise, değişkenler bağımsızdır.

Örneğin "RA ile kontrol grubu arasında eğitim durumu açısından fark yoktur", "oral ve parenteral demir preparatlarının allerjik yan etki açısından farkı yoktur" denilebilir.

Ancak ki-kare testi önemli fark olduğu sonucunu verirse, yani $p<0.05$ bulunursa, değişkenlerin bağımlı olduğu söylenebilir, ama bağımlı değişken ordinalse, bu değişkeninin düzeyleri ile bağımsız değişkenin düzeyleri arasındaki ilişkinin yönü hakkında bilgi vermez. Örneğin Tablo 1'de ki-kare testi ile gruplar arasında önemli fark olduğu sonucuna varıldığını varsayıyalım. Buna dayanarak yalnızca "RA ile kontrol grubu arasında eğitim durumu açısından fark vardır" denebilir.

"Eğitim durumu kötüden iyiye gittikçe RA ile kontrol grubu arasındaki fark değişiyor mu, değişiyorsa hangi yönde değişiyor, artma ya da azalma yönünde mi, hangi eğitim düzeyinde RA ile kontrol grubuna ait deneklerin oranları anlamlı olarak farklı?". Bu soruların hiçbirine ki-kare testi ile karar verilemez. Çünkü ki-kare testi, değişkenlerin bağımlı olup olmadıklarını test eder, ya da değişkenler arasındaki ilişkiyi genel olarak değerlendirdir.

2x2'lik tablolar için ki-kare düzeltmesi

Bazen 2X2'lik çapraz tablolardan ki-kare testi ile elde edilen ki-kare değerinde, güvenilirliği artttırmak için devamlık düzeltmesi (*continuity correction*) ya da Yates' düzeltmesi denilen bir hesaplama ile değişiklik yapılması önerilmektedir. Bu durumda düzeltme yapılmadan hesaplanan değerden daha küçük bir ki-kare değeri elde edilir. Ki-kare değeri küçülünce, p değeri büyür ve iki değişken arasında ilişki bulmak zorlaşır. Örneğin düzeltmesiz yapıldığında sınırdı önemli bulunan bir p değeri (örneğin 0.04), Yates düzeltmesi yapıldığında önemsiz bir değere değişimdir (örneğin 0.07 olabilir).

2X2'lik çapraz tablolarda beklenen değerlerden birinin 2'den küçük olması durumunda Fisher kesin ki-kare testi uygulanması bir zorunluluktur, ama Yates' düzeltmesini uygulayıp uygulamamak yalnızca bir tercih sorunudur.

İKİDEN FAZLA BAĞIMSIZ GRUP ORANININ KARŞILAŞTIRILMASI

2x2'den büyük çapraz tablolar için ki-kare testi

Ki-kare testi yalnız 2X2'lik çapraz tablolara uygulanmaz. Daha çok gözlü çapraz tablolarda da iki nominal değişken arasında ilişki olup olmadığı ki-kare

re testi ile incelenebilir. Formülü 2X2'lik çapraz tablolar için kullanılan formülle aynıdır. Sıra sayısı r ile, sütun sayısı c ile gösterildiğinde, ki-karenin hesaplanmasında $r \times c$ adet terim bulunur.

ÖRNEK 58

Örnek 2'deki çalışmada 50 kişilik başka bir gruba da nitrat verildiğini ve Tablo 7'deki sonuçların elde edildiğini varsayıyalım.

Tablo 7 Örnek 58'deki çalışmanın sonuçlarının gösterildiği çapraz tablo.

	Öldü	Yaşıyor	Toplam
Beta-bloker	6	44	50
Kalsiyum antagonisti	12	38	50
Nitrat	15	35	50
Toplam	33	117	150

Bu tabloda 3 sıra, 2 sütun olduğu için 2X3'lük bir çapraz tablodur. Beklenen sayılar "Öldü" sütunundaki gözler için 11, "Yaşıyor" sütunundaki gözler için 39'dur. Gözlerin hepsinde, hesaplandığında beklenen sayıların 5'ten büyük olduğu için görülecektir. Bu nedenle ki-kare testi yapılabilir. Hesaplanan ki-kare değeri, tablodaki ki-kare değerinden küçükse (yani bulunan p değeri >0.05 ise), ilaç tipi ile ölüm oranı arasında ilişki yoktur. Hesaplanan ki-kare değeri, tablodaki ki-kare değerinden büyükse (yani bulunan p değeri <0.05 ise), ilaç tipi ile ölüm oranı arasında ilişki olduğu söylenebilir. Ancak bu durumda "Hangi ilaç grubunda ölüm oranı diğerlerinden farklıdır?" sorusu henüz açıklığa kavuşmamıştır.

P değeri <0.05 ise ne yapmalı?

Bu sorunun net bir yanıtı yoktur. Bir yöntem olarak, her sıradaki ki-kare değerlerine bakılarak, en büyük ki-kare değerine sahip olan sıra çıkarılır ve kalan sıra ve sütunlarla ki-kare testi tekrarlanır. Bu kez ki-kare testi ile bulunan ilişkinin derecesi ömensiz olursa, yani $p > 0.05$ ise, çıkarılmış olan sırayı belirleyen ilaç grubunda ölüm oranının diğer gruplardan farklı olduğu, diğer iki grup arasında ise fark olmadığı anlaşılır.

Örneğin beta-bloker sırasındaki iki gözdeki ki-kare terimlerinin toplamı bu sıranın ki-kare değerini verir. Üç sıra içinde en büyük ki-kare değeri bu

gruba aitse, bu grup çıkarılır ve diğer iki gruptaki denekleri içeren 2X2'lik yeni bir çapraz tablo yapılır ve ölüm oranları karşılaştırılır. İkisi arasında fark yoksa "beta-bloker kullananlarda ölüm oranının diğer iki gruba göre düşük olduğu, kalsiyum antagonist ile nitrat kullananlarda ise ölüm oranının farklı olmadığı" sonucuna varılır.

2X2'den büyük çapraz tablolarda iki değişken arasında ilişki bulunduğuunda, en büyük ki-kare değerine sahip olan sıra ya da sütun çıkarıldıkten sonra kalan sıra ve sütunlar için ki-kare testinin tekrarlanmasıında fark ya da ilişki öneemsiz bulunduğuunda, yani $p>0.05$ bulunduğuunda son verilir.

Bu yöntemin önemli bir sakincası, testin tip I hatasını arttırmasıdır. Yani arasında gerçekte fark olmayan grup oranlarının farklı bulunması gibi bir yanlışlık yol açabilir. Bu nedenle mümkün olduğu kadar kullanılmamalıdır.

Beklenen sayılar küçükse ne yapmalı?

2X2'den büyük çapraz tablolarda ki-kare testi kullanılabilmesi için de, gözlerin hepsindeki beklenen sayıların 2 ya da daha büyük ve yarısı ya da daha fazlasındaki beklenen sayıların 5 ya da daha büyük olması gereklidir. Fisher testi yalnız 2X2'lik çapraz tablolarda kullanıldığı için, bu koşulun sağlanmadığı durumlarda sıra ve sütunu oluşturan grupların birleştirilmesi yoluna gidilmelidir.

ÖRNEK 59

Eğitim durumu ile stress arasında ilişki olup olmadığını araştırıldığı bir çalışmada farklı eğitim düzeyindeki 100 kişiye anket uygulanıyor. Stress derecesi "minimal", "orta derecede" ve "aşırı" olarak ordinal skalada değerlendiriliyor. Çalışmanın sonuçları Tablo 8'de verilmiştir.

Tablo 8 Örnek 59'daki çalışmanın sonuçlarının gösterdiği çapraz tablo. G sütunlarında gözlenen, B sütunlarında beklenen sayılar verilmiştir. Beşten küçük olan beklenen sayılar italik harflerle gösterilmiştir.

	Stress'in derecesi						Toplam	
	Minimal		Orta derecede		Aşırı			
	G	B	G	B	G	B		
İlkokul	6	2	1	4	1	2	8	
Ortaokul	6	4.5	7	9	5	4.5	18	
Lise	12	16	41	32	11	16	64	
Üniversite	1	2.5	1	5	8	2.5	10	
Toplam	25		50		25		100	

Görüldüğü gibi tablo 4X3, yani 12 gözlüdür. Tabloda G sütunlarında gözlenen, B sütunlarında beklenen sayılar verilmiştir. Oniki gözden 7'sinde beklenen sayılar 5'ten küçük olduğu için ki-kare testi yapılamaz. Bu nedenle ya sıra, ya da sütun değişkeninden birinde gruplar birleştirilmelidir. Sıra değişkeninde gruplar birleştirilirse, Tablo 9'daki çapraz tablo elde edilecektir.

Tablo 9 Tablo 8'deki eğitim durumu kategorilerinde birleştirme yapıldığında elde edilen çapraz tablo. G sütunlarında gözlenen, B sütunlarında beklenen sayılar verilmiştir. Beşten küçük olan beklenen sayılar italik harflerle gösterilmiştir.

	Stress'in derecesi						Toplam
	Minimal		Orta derecede		Aşırı		
	G	B	G	B	G	B	
İlkokul ve Ortaokul	12	6.5	8	13	6	6.5	26
Lise	12	16	41	32	11	16	64
Üniversite	1	2.5	1	5	8	2.5	10
Toplam	25		50		25		100

Görüldüğü gibi "İlkokul" ve "Ortaokul" grupları birleştirildiğinde 9 gözden yalnız ikisinde beklenen sayılar 5'ten küçük olduğu için artık ki-kare testi uygulanabilir.

Eğitim durumu yerine stress derecesi grupları da birleştirilebilirdi. "Minimal stress" ve "Orta derecede stress" grupları birleştirildiğinde Tablo 10'daki çapraz tablo elde edilecektir.

Tablo 10 Tablo 8'deki stress kategorilerinde birleştirme yapıldığında elde edilen çapraz tablo. G sütunlarında gözlenen, B sütunlarında beklenen sayılar verilmiştir. Beşten küçük olan beklenen sayılar italik harflerle gösterilmiştir.

	Stress'in derecesi				Toplam	
	Minimal ya da Orta derecede		Aşırı			
	G	B	G	B		
İlkokul	7	6	1	2	8	
Ortaokul	13	13.5	5	4.5	18	
Lise	53	48	11	16	64	
Üniversite	2	7.5	8	2.5	10	
Toplam	75		25		100	

Sekiz gözden üçünde, yani yarısından azında beklenen sayı 5'ten küçük olduğu için ki-kare testi uygulanabilir. Sıra ya da sütun değişkenlerinden hangisinin kategorilerinin birleştirileceğinin istatistik açıdan önemi yoktur. Klinik açıdan hangisi önemli ise, diğer değişkenin kategorize edilmesi daha yararlı olacaktır. Yukarıdaki örnekte eğitim durumunun çeşitli değişkenler üzerindeki etkileri araştırılıyor ve bu değişkenlerden biri de stress ise, önemli olan değişken "eğitim durumu"dur. Bu nedenle "stress" kategorilerinin birleştirilmesi amaca daha uygundur. Stress oluşumuna etki eden etkenlerin araştırıldığı bir çalışmada ise önemli olan stress olduğu için "eğitim durumu" kategorilerinin birleştirilmesi yoluna gidilmelidir.

Kategoriler birleştirilip, çapraz tablo küçüldüğü halde, hala gözlerdeki beklenen sayılar küçükse, kategori birleştirme işlemine devam edilmelidir. Örneğin "eğitim durumu" kategorileri "İlkokul ve Ortaokul", "Lise" ve "Üniversite" olarak birleştirildiği halde, gözlerin yarısından çoğu beklenen sayılar 5'ten küçükse, "İlkokul ve Ortaokul" ve "Lise ve Üniversite" olarak kategoriler birleştirilebilir. 2X2'lik çapraz tablo oluşana kadar kategori birleştirmeye devam edildiği halde, beklenen değerler küçük kalıyorsa, Fisher testine geçilmelidir.

Relatif risk (risk oranı)

Bir olayın ortaya çıkması ile ilişkili değişkenleri belirlemek bir çok çalışmaının asıl konusudur. Bir etkenle bir olay arasındaki ilişkisinin derecesini ölçen, sık kullanılan bir indeks relatif risk ya da risk oranıdır. Risk etkeninin var olduğu ve olmadığı durumlarda olayın sıklığının oranı olarak hesaplanır.

Kısaca RR olarak gösterilen risk oranı, kohort çalışmalarında kullanılan bir

istatistik yöntemdir. Bilindiği gibi, kohort çalışmalarında risk etkeni olan ve olmayan sağlıklı deneklerden oluşan iki grupta denek izlemi başlar ve belirli bir süre sonra her iki grupta da kaç denekte hastalığın ortaya çıktıgı saptanır.

RR, risk etkeni olan grupta hastalığın ortaya çıkma insidansının, risk etkeni olmayan grupta hastalığın ortaya çıkma insidansına oranıdır. RR, 1'den ne kadar büyükse, risk etkeninin hastalığın ortaya çıkmasındaki rolü o kadar fazladır. Tersine RR, 1'den küçükse, "risk etkeni grubunda hastalık daha az sıklıkla ortaya çıkmaktadır ya da bu etken hastalığın ortaya çıkışını önlemektedir" sonucuna varılabilir. Kohort çalışmalarında, risk etkenlerinin belirlenmesi dışında bir ilaçın yararlı etkisi de araştırılabılır.

ÖRNEK 60

Miyocard infarktüsü sonrası ürokinaz kullanılmasının mortaliteyi azaltmadan rolü olup olmadığını araştırmak için yapılan bir kohort çalışmada bir grup hastaya ürokinaz, bir grup hastaya plasebo veriliyor ve tüm hastalar bir yıl süreyle izleniyor.

Bu çalışmada "risk etkeni grubu"nu ürokinaz alanlar oluşturmaktadır. (Göründüğü gibi risk etkeni, her zaman zararlı bir etken değildir. Asıl araştırma konusu olan etken, risk etkeni olarak tanımlanır.) Bu grubun bir yıllık ölüm insidansının, plasebo grubunun ölüm insidansına oram, RR'i verir. Burada beklenen, RR'in 1'den küçük olmasıdır.

Tablo 11'de görüldüğü gibi "Risk etkeni" grubunda ölüm insidansı $98/945 = \%10.4$, plasebo grubunda ise $152/939 = \%16.2$ 'dir. Birincinin ikinciye oranı olan RR ise 0.641'dir. $RR < 1$ olduğu için, ürokinazın ölüm insidansını azalttığı anlaşılmaktadır.

Tablo 11 Örnek 60'daki çalışmanın sonuçlarının gösterildiği çapraz tablo.

	Öldü	Yaşıyor	Toplam
Ürokinaz	98	847	945
Plasebo	152	787	939

Sigara ile kalp hastalığı arasında ilişki olup olmadığını araştırmak için kalp hastası olmayan 2000 kişi 15 yıl süreyle izleniyor. Bu kişilerin 1000 tanesi sigara içiyor, 1000 tanesi içmiyor. 15 yıl sonunda her iki grupta kaç kişide kalp hastalığı ortaya çıktıgına baklıyor. Tablo 12'de sonuçlar verilmiştir.

Tablo 12 Örnek 61'deki çalışmanın sonuçlarının gösterildiği çapraz tablosu.

		Kalp hastalığı		
Sigara		Var	Yok	Toplam
İçiyor		100	900	1000 (%50)
İçmiyor		50	950	1000 (%50)
Toplam		150 (%7.5)	1850 (%92.5)	2000 (%100)

Göründüğü gibi 15 yıllık kalp hastalığı insidansı, sigara içenlerde 100/1000, sigara içmeyenlerde 50/1000. Risk oranı ise $(100/1000) / (50/1000) = 2$.

Yani sigara içenlerde kalp hastalığı gelişmesi riski içmeyenlere göre iki kat daha fazladır. Risk oranının %95 güven sınırları, yani %95 olasılıkla alabileceği minimum ve maksimum değerler ise 1.44 ve 2.78 olarak hesaplanmıştır. Yani %5 yanılma payıyla sigara içenlerde kalp hastalığı riski, sigara içmeyenlere oranla en az 1.44, en çok 2.78 kat artmıştır.

Eğer alt güven sınırı 1'den büyükse, (ya da üst güven sınırı 1'den küçükse), yani %95 güven aralığı 1'i içermiyorsa, $p < 0.05$ önem derecesiyle belirlenen etken, olayın ortaya çıkması riskini arttırıyor (ya da azaltıyor) demektir.

Bu örnekte 15 yıl sigara içmek, kalp hastalığı riskini en az 1.44 kat arttıguna göre, kalp hastalığı riskini istatistiksel olarak önemli derecede artırıyor demektir.

Odds oranı

Kısaca OR olarak gösterilen "odds" oranı, retrospektif (*case-control*) çalışma malarında kullanılan bir istatistik terimidir. Bilindiği gibi, bu tip çalışmalarında hastalıklı ve sağlıklı deneklerden oluşan iki grupta ilgilenilen risk etkeninin var olup olmadığı sorgulanır ya da araştırılır.

Örnek 61'deki çalışmada, kalp hastalığı olmayan kişiler alınıp, kalp hastalığı ortaya çıkma insidansını, sigara içmenin arttııp arttırmadığını araştırmıştık. Retrospektif çalışmalarda ise, bunun aksine hastalığı olan ve olmayan iki grup alıp, bu iki grup arasında riski artttığı kestirilen etken açısından fark olup olmadığı araştırılmıştır.

ÖRNEK 62

Sigara ile kalp hastalığı arasında ilişki olup olmadığını araştırmak için 200 koroner kalp hastası ve 200 normal kişi alınıyor ve sigara içip içmedikleri belirleniyor. Tablo 13'de sonuçlar verilmiştir.

Tablo 13 Örnek 62'deki çalışmanın sonuçlarının gösterildiği çapraz tablo.

Sayı (Satır yüzdesi)	Sigara		Toplam
	İçiyor	İçmiyor	
Kalp hastalığı var	30 (%30)	70 (%70)	100 (%50)
Kalp hastalığı yok	10 (%10)	90 (%90)	100 (%50)
Toplam	40 (%20)	160 (%80)	200 (%100)

Bu çalışma retrospektif olduğu için insidans ve risk oranı hesaplanamaz. Bunun yerine *odds* oranı hesaplanabilir.

Odds, her grupta risk etkeni olanların sayısının olmayanlara oranıdır.

Örneğin kalp hastaları grubunda sigara içenlerin içmeyenlere oranına "Kalp hastaları grubunun risk etkenine maruz kalma *odds*" u adı verilir. *Odds* terimi, İngilizce'de olasılık anlamına gelen bir terimdir. Ancak olasılık tanımından farklı olarak, risk etkenine sahip olan deneklerin sayısının toplam denek sayısına değil, risk etkenine sahip olmayan deneklerin sayısına oranı olarak tanımlanır. Bu nedenle Türkçe'de tam bir karşılığı bulunmayan *odds* terimi, İngilizce'den çevrilmeden kullanılacaktır.

Yukarıdaki örnekte sigara içme *odds*'u, kalp hastalığı grubunda 30/70, kontrol grubunda 10/90'dır. *Odds* oranı ise hasta grubun *odds*'unun, kontrol grubun *odds*'una orandır $((30/70) / (10/90) = 3.85$). *Odds* oranının %95 güven

sınırları 1.77 ve 8.42 olarak hesaplanmıştır. Bu örnekte sigara içme *odds*'u, kalp hastalığı olanlarda en az 1.77 kat fazla olduğuna göre, kalp hastalığı olanlarda, istatistiksel olarak önemli derecede daha fazla kişi sigara içmiştir.

Risk oranı mı, *odds* oranı mı?

ÖRNEK 63

Miyokard infarktüsü (MI) geçirenlerde sigaranın risk etkeni olup olmadığı araştırılıyor. Bu amaçla yapılan bir çalışmada miyokard infarktüsü geçiren deneklerden bir grup ve sağlıklı kişilerden bir grup olmak üzere iki grupta sigara içme öyküsü soruluyor ve Tablo 14'deki sonuçlar elde ediliyor.

Tablo 14 Örnek 63'teki çalışmanın sonuçlarının gösterildiği çapraz tablo.

	MI	Sağlıklı
Sigara içiyor	170	90
Sigara içmiyor	30	110
Toplam	200	200

MI grubunun *odds*'u $170/30 = 5.667$, kontrol grubunun *odds*'u $90/110 = 0.818$ olarak hesaplanır. *Odds*'ların oranı $5.667/0.818 = 6.93$ 'tür. Yani MI grubundaki bir kişinin, kontrol grubundaki bir kişiye göre sigaraya maruz kalma *odds*'u yaklaşık 7 kat daha fazladır. OR, 1'den ne kadar büyükse, risk etkeninin hastalığın ortaya çıkmasındaki rolü o kadar fazladır. Tersine OR, 1'den küçükse, "sağlıklı gruba göre hasta grupta risk etkenine daha az sıklıkla rastlanmaktadır" sonucuna varılabilir.

RR ile OR arasında seçim yaparken, çalışmanın tipini belirlemek gerekir. Kohort çalışmalarında RR, retrospektif çalışmalarında OR kullanılır.

İkisi arasındaki farkı görmek için bir önceki çalışmanın prospektif kohort olarak yapıldığını düşünün. Yani sigara içen 260 sağlıklı kişi ile içmeyen 110 sağlıklı kişinin izleme alındığını, ve 5 yıl sonra birinci grupta 170, ikinci grupta 30 kişinin MI geçirdiğini varsayıyalım. Aynı çapraz tablo elde edilecektir (Tablo 15). Ama kohort tipinde bir çalışma olduğu için OR yerine RR kullanılmalıdır.

Tablo 15 Örnek 63'teki çalışma prospektif olsaydı, sonuçların gösterileceği çapraz tablo.

Sigara	Kalp hastalığı		Toplam
	Var	Yok	
İçiyor	170	90	260
İçmiyor	30	110	140

Sigara içenler grubunda MI insidansı $170/260 = \%65$ ve sigara içmeyenler grubunda MI insidansı $30/140 = \%21$ 'dir. İkisinin oranı olan RR ise 3.05'tir. Aynı tablo için hesaplanan OR 6.93, RR ise 3.05'tir. İkisi birbirinden çok farklıdır. Çünkü ifade ettikleri şey de farklıdır. OR'ye bakarak anlaşılan şudur: MI geçirmiş olan bir kişinin sağlıklı bir kişiyle karşılaşıldığında sigara içmiş olma şansı (olasılığ = *odds'u*) 7 kat daha fazladır. RR'na bakarak anlaşılan ise şudur: Sigara içmek MI riskini 3 kat artttırmaktadır.

Risk oranı mı, ki-kare mi?

2X2'lik bir çapraz tabloya Z testi ve ki-kare testi uygulanması dışında, risk oranı ya da odds oranı da hesaplanabilir. Bunların herbiri, bazı kısıtlılıklarla diğerinin yerine geçebilir. Eğer ilgilenilen konu yalnızca iki oran arasında fark olup olmadığı ise Z testi ya da ki-kare testi kullanılabilir. Eğer değişkenlerden biri, diğer değişken açısından risk etkeni ise ve çalışma düzeni kohort tipteyse risk oranı, çalışma düzeni retrospektifse *odds* oranı kullanılabilir. Bu testlerin birisinde iki grup oranı arasındaki fark ya da iki değişken arasındaki ilişkili önemli bulunursa, diğer testlerde de önemli bulunur. Bu nedenle, tıp literatüründe giderek daha sıklıkla ki-kare ya da Z testi yerine risk oranı ya da *odds* oranı kullanılmıştır.

BAĞIMLI ORANLARIN KARŞILAŞTIRILMASI

McNemar testi

ÖRNEK 64

Kan şekeri üzerine etkisini araştırmak için 100 diyabetik hastaya yeni bir diüretiğin verildiğini ve önce ve 3 ay sonra açlık kan şekeri (AKŞ) ölçüldüğünü varsayılmı-

Diüretiğin kan şekerine etkisini araştırmanın bir yolu tedavi öncesi ve sonrası AKŞ ortalamalarının bağımlı gruptarda Student t testi ile karşılaştırılmasıdır. AKŞ'ının bir hastada 80 mg/dL'den 105 mg/dL'ye çıkışının önemli olmadığını varsayıarak, kan şekeri ≥ 120 mg/dL olanların oranının tedavi ile ne kadar değiştiği incelenmek istiyorsa, bir ki-kare testi uyarlaması olan McNemar testi uygulanır. Ki-kare testi yapılamaz, çünkü tedavi öncesi ve sonrası gruplar aynı denekleri içерdiği için bağımlıdır. Ki-kare testinin uygulanabilmesi için grupların bağımsız olması gereklidir (Tablo 16).

Tablo 16 Örnek 64'teki çalışmanın sonuçlarının gösterildiği çapraz tablo.

		Tedavi sonrası		Toplam
Tedavi öncesi	AKŞ<120	AKŞ<120	AKŞ≥120	
	AKŞ≥120	c = 10	d = 30	40
Toplam		30	70	100

Tedavi öncesi hastaların %40'ında KŞ ≥ 120 iken, tedaviden sonra %70'ye yükselmiştir. İki oran arasındaki farkın önemli olup olmadığını söylemek için hesaplanan ki-kare değeri, serbestlik derecesi=1'e karşılık gelen tablo değeri ile karşılaştırılmalıdır.

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c}$$

Bu formüle göre hesaplandığında ki-kare değeri 16.82'dir. 0.05 yanılma düzeyinde ve 1 serbestlik derecesindeki tablo değeri olan 3.84'ten büyük olduğu için iki oran arasındaki farkın önemli olduğu sonucu çıkmaktadır. McNemar testi, kişiler ya da tanısal yöntemler arasındaki uyumu test etmek için de kullanılabilir.

ÖRNEK 65

İki patolog, 100 akciğer biyopsi hazırlatmasını, diğerinden habersiz olarak "benign" ya da "malign" olarak rapor ediyor. Ne derece uyumlu olduklarını test etmek için çapraz tabloyu oluşturalım (Tablo 17).

Tablo 17 Örnek 65'teki çalışmanın sonuçlarının gösterildiği çapraz tablo.

		Dr. Ahmet		Toplam
		Malign	Benign	
Dr. Mehmet	Malign	72	8	80
	Benign	5	15	20
	Toplam	77	23	100

Dr. Mehmet, Dr. Ahmet'in "malign" olarak rapor ettiği preparatlardan 5'ine "benign", Dr. Ahmet'in "benign" olarak rapor ettiği preparatlardan 8'ine ise "malign" demiş. Yani 13 / 120 preparattaki uyumsuzluk önemli derecede midir? McNemar testi ile hesaplanan p değeri 0.58'dir. Yani iki patolog arasındaki fark önemli değildir ya da ikisi arasında önemli derecede uyumsuzluk yoktur. Uyuman derecesi ise K (kappa) istatistiği ile test edilir. McNemar testinde iki kişi/yöntem uyumu test edilirken, hangisinin daha doğru olduğu konusunda bir ön bilgi gerekmekz. Dr. Ahmet, akciğer patolojisi konusunda uzman bir patolog ise (altın standart), Dr. Mehmet'in bu konuda ne derece başarılı olduğu McNemar testi ya da kappa istatistiği yerine duyarlılık (sensitivity) ve özgürlük (spesifisite) analizi ile araştırılmalıdır (Bölüm 17).

Cochran's Q testi

Eğer karşılaştırılacak bağımlı dikotom değişken sayısı 2 ise uygulanması gereken test McNemar testi iken, bağımlı dikotom değişken sayısı 2'den çoksa McNemar testinin bir varyantı olan Cochran Q testi uygulanmalıdır. Örnek

64'teki patolog sayısı 2'den çok olsaydı, McNemar testi yerine Cochran Q testi uygulanması gerekecekti.

Marjinal homojenlik testi

McNemar testinin bir varyantı da marjinal homojenlik testidir. Eğer ölçülen değişken ordinalse McNemar testi yerine marjinal homojenlik testi uygulanmalıdır. Örnek 64'te hastalar AKŞ düzeyine göre 2'den çok gruba ayrılsaydı ya da Örnek 65'teki iki patoloğun değerlendirilmesi 2'den çok düzeyli olsaydı, McNemar testi yerine marjinal homojenlik testi uygulanması gerekecekti.

Kappa istatistiği

Uyum ölçümü (*measurement of agreement*) adı verilen bu teknik, aynı grup denek üzerinde iki farklı gözlemeçinin değerlendirmelerinin uyumunu test etmekte kullanılır.

ÖRNEK 66

İki doktor, 72 hastayı muayene ederek, "splenomegali var, şüpheli ya da yok" diye rapor ediyorlar. Sonuçlar Tablo 18'de verilmiştir. İki doktorun muayenesi uyumlu mudur?

İki doktor 17 hastada splenomegali var, 13 hastada yok demişler. 12 hasta da ise her ikisi de "şüpheli" diye rapor etmişler. Yani 72 hastanın 42'sinde (%58) aynı, diğer hastalarda ise yorumları farklı. Bu 42 hastada (%58) rastlantısal sonuçlar da olabilir. Hesaplanan, tesadüfü olmayan uyum Cohen kappa (K) değeri'dir. Bu örnekte K değeri 0.362 olarak hesaplanmıştır, yani tesadüfü olmayan uyum %36'dır. K değeri -1 ile 1 arasındadır. 0 ise iki gözlemeçi arasında hiç uyum yoktur, 1 ise uyum çok mükemmeldir, -1 ise birinin "ak" dediği ne diğeri "kara" demektedir.

Tablo 18 Dalak muayenesinde iki doktorun uyumu ile ilgili çalışmanın çapraz tablosu.

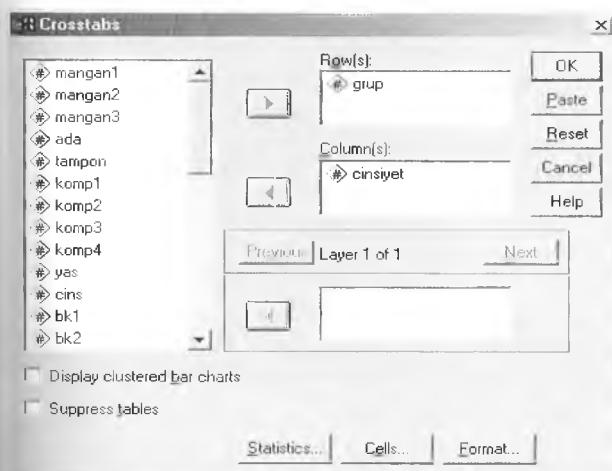
Dr. Mehmet	Sayı Toplam yüzde	Dr. Ahmet			Toplam
		Var	?	Yok	
Var	17	4	8	29	
	23.6	5.6	11.1	40.3	
	5	12	0	17	
?	6.9	16.7	0.0	23.6	
	10	3	13	26	
Yok	13.9	4.2	18.1	36.1	
	32	19	21	72	
Toplam	44.4	26.4	29.2	100.0	

SPSS'TE BAĞIMSIZ GRUP ORANLARI NASIL KARŞILAŞTIRILIR?

Crosstabs menüsü

Analyze >> Descriptive Statistics >> Crosstabs

Yukarıdaki seçenekle Crosstabs diyalog kutusu açılır (Şekil 3). Dosyadaki sayısal ve kısa string değişkenlerin adı listelenir. Satır (row) ve sütun (column) değişkenleri kutularına bu değişkenlerden birden çok seçilebilir.



Şekil 3
SPSS'te "Descriptive Statistics >> Crosstabs" seçenekinden "Crosstabs" diyalog kutusu.

Her satır ve sütun değişkeni çifti için ayrı çapraz tablolar yapılır. Örneğin satır değişkeni olarak 4, sütün değişkeni olarak 3 değişken seçilmişse, toplam 12 adet çapraz tablo yapılır.

Ayrıca tabaka (*layer*) değişkeni kutusuna da değişkenlerden bir ya da bir kaç tane seçilebilir. Örneğin cinsiyet tabaka değişkeni olarak seçilmişse, erkekler için 12, kadınlar için 12 olmak üzere toplam 24 çapraz tablo yapılacaktır. Tabaka değişkenleri, 8 içe içe alt tabaka oluşturacak şekilde de seçilebilir.

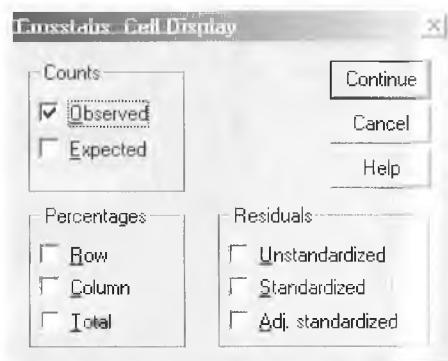
Örneğin grup değişkeninin "tip I DM, tip II DM ve kontrol grubu" değerleri aldığıını düşünürsek, önce grup olarak tabakalandırıp, sonra cinsiyet değişkenini alt tabaka değişkeni olarak eklersek, her hastalık grubu ve kontrol grubu için erkek ve kadınlar olmak üzere çapraz tablolar yapılacaktır.

İstatistik ve satır ve sütun yüzdeleri içermeyen çapraz tablolar yapılması önceden seçilmiştir. Eğer *Statistics* seçeneğinden istatistiklerle ilgilenip, çapraz tabloların yapılmasını istemiyorsanız, *Suppress tables*'ı seçin (Önceden seçilmemiş durumdadır).

Crosstabs diyalog kutusundan *Statistics*, *Cells* ve *Format* altmenülerine geçilebilir.

Cells altmenüsü

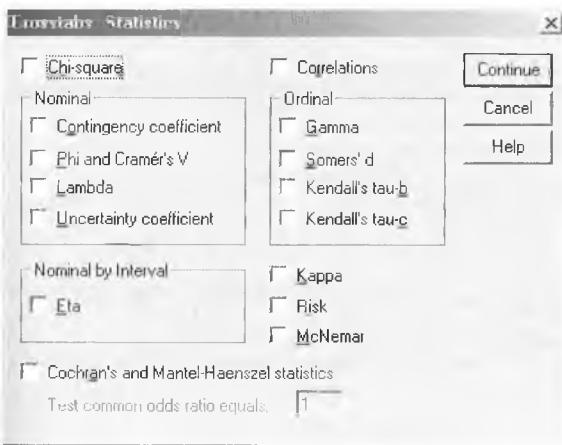
Cells altmenüsünden gözlenen (*observed*) ve beklenen (*expected*) sayılar, satır, sütun ve toplam yüzdeler ve rezidüller seçilebilir (Şekil 4). Gözlenen sayılar önceden seçilmiştir. Çapraz tabloda gözlenen sayılar dışında beklenen sayılar ya da yüzdelerin de yer olmasını istiyorsanız uygun seçimi yapmalısınız. Rezidüllerden burada söz edilmeyecektir.



Şekil 4
SPSS'te "Crosstabs" diyalog kutusundan ulaşılan "Cell Display" altmenüsü.

Statistics altmenüsü

Statistics altmenüsündeki seçeneklerin hiç biri önceden seçilmemiştir (Şekil 5). Yani bu menüye girmezseniz, istatistik yapılmadan yalnız çapraz tabloları çizdirirsınız. Statistics menüsünde nominal ve ordinal verilere uygulanaabilecek çapraz tablo istatistiklerine ilişkin çok sayıda seçenek yer almaktadır.



Şekil 5
SPSS'te "Crosstabs" diyalog kutusundan ulaşlan "Statistics" menüsü.

Ki-kare (χ^2) (chi-square)

Bu seçenek seçildiğinde tabloya ait Pearson χ^2 , likelihood ratio χ^2 ve Mantel-Haenszel doğrusal ilişki χ^2 testleri yapılır. En sıkılıkla kullanılan değer Bölüm 13'te söz edilen Pearson χ^2 testine ait olandır.

Şekil 6'da *Linear-by-Linear Association* satırında yer alan Mantel-Haenszel doğrusal ilişki χ^2 testi değişkenlerin ordinal olduğu durumlarda geçerlidir. 2X2'lik çapraz tablolar için Yates düzeltmesi ve gerekirse Fisher testi de yapılır. Örneğin "tip I DM ve tip II DM" olarak seçilen iki grup hastanın cinsiyete göre çapraz tablosu ve χ^2 testi sonuçları Şekil 6'da gösterilmiştir.

Cinsiyet * Grup Crosstabulation					
Cinsiyet	Erkek		Grup		Total
			Tip I DM	Tip II DM	
Cinsiyet	Erkek	Count	11	8	19
		Expected Count	9.3	9.7	19.0
Kadın		Count	9	13	22
		Expected Count	10.7	11.3	22.0
	Total	Count	20	21	41
		Expected Count	20.0	21.0	41.0

Şekil 6
Tip I DM ve tip II DM gruplarındaki erkek ve kadınların sayılarını gösteren çapraz tablo ve χ^2 istatistikleri. Satır değişkeni "cinsiyet", sütun değişkeni "grup" olarak seçilmiştir.

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.177 ^b	1	.278		
Continuity Correction ^a	.596	1	.440		
Likelihood Ratio	1.183	1	.277		
Fisher's Exact Test				.354	.220
Linear-by-Linear Association	1.149	1	.284		
N of Valid Cases	41				

a. Computed only for a 2x2 table
b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 9.27.

Value sütünunda χ^2 değerleri, Sig sütünunda karşılık gelen p değerleri verilmiştir. p değeri <0.05 olmadığı için ($p=0.28$) iki gruptaki deneklerin erkek/kadın oranlarının farklı olmadığı anlaşılmaktadır. Continuity correction satırındaki değerler, Yates düzeltmesi yapıldıktan sonra elde edilen değerlerdir. Gözlerde beklenen değeri 5'den küçük olan olmadığı için Fisher testi uygulanması gereklidir.

Fisher testi ile hesaplanan p değeri tek-yönlü ve iki-yönlü olarak iki kez belirtilemiştir. Eğer belli bir cinsiyetin bir grupta daha fazla olup olmadığı gibi belli bir yönde hipotezi kurmuşsak, p değeri olarak 0.22'yi almak daha uygun olacaktır. Genellikle istatistik analize geçmeden önce, "Şu değişken A grubunda B grubundan daha sıktır" gibi tek yönlü (one-tailed) yerine "Şu değişkenin dağılımı A ve B grupları arasında farklıdır" gibi iki yönlü (two-tailed) varyanslar yaptığımız için iki yönlü p değerlerini kullanmak daha uygun olur.

Risk

Yalnız 2X2'lik tablolar için risk oranı ve odds oranı elde edilir. Risk oranı için risk etkeni satır değişkeni olarak, hastalık grubu sütun değişkeni olarak seçilmelidir.

Odds oranı için satır ve sütun değişkeninin seçim sırası önemli değildir. Stress, RA gelişmesi için bir risk etkeni midir? Şekil 7'deki tabloda risk oranları ve *odds* oranı verilmiştir.

Stress * GRUP Crosstabulation					
Stress	Var	Count	GRUP		Total
			RA	Kontrol	
Yok	Count	11		8	19
	% within Stress	57.9%		42.1%	100.0%
Total	Count	30		15	45
	% within Stress	66.7%		33.3%	100.0%

Risk Estimate				
	Value	95% Confidence Interval		
		Lower	Upper	
Odds Ratio for Stress (Var / Yok)	1.974	.562	6.939	
For cohort GRUP = RA	1.262	.806	1.977	
For cohort GRUP = Kontrol	.639	.280	1.458	
N of Valid Cases	45			

Şekil 7
Romatoid artrit-stress risk ilişkisinin çapraz tablosu.

Eğer çalışma sağlıklı kişilerin aldığı ve stressi olanlarda ve olmayanlar da ne kadar sıklıkla RA geliştiğinin izlendiği bir çalışmaya (prospektif), risk oranlarının kullanılması gereklidir. Bu nedenle risk etkeni (stress) satır değişkeni, hastalık grubu (RA) sütun değişkeni olarak girilmiştir. *For cohort grup = RA* satırındaki değerler risk oranı ve güven sınırlarını vermektedir. RA riski, stressi olanlarda olmayanlara göre 1.26 kat fazladır; ama alt güven sınırı <1 (0.81) olduğu için, yani güven aralığı (0.81-1.98) 1'i içerdigi için "stress, RA riskini istatistiksel olarak önemli ölçüde artırmamaktadır" denebilir.

Eğer çalışma retrospektifse, yani RA'lı ve normal kişiler alınıp stress olup olmamasına göre grupperlendirilmişlarsa, o zaman risk oranı değil, odds oranı

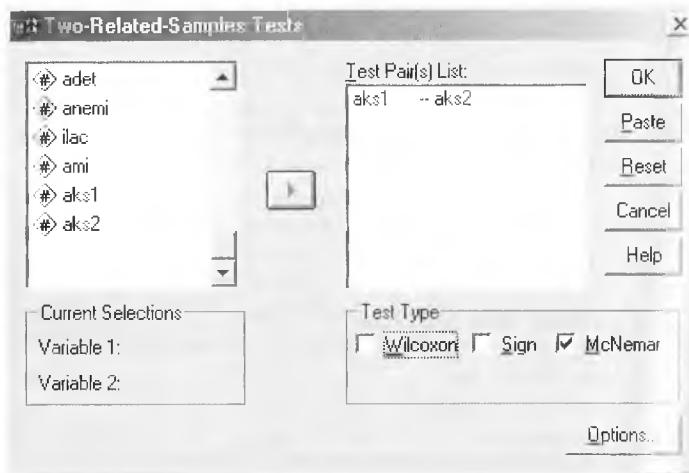
kullanılmalıdır. *Odds ratio for Stress (Var / Yok)* satırında verilen değerler *odds* oranı ve güven sınırlarıdır.

SPSS'TE BAĞIMLI GRUP ORANLARI NASIL KARŞILAŞTIRILIR?

McNemar testi

Analyze >> Descriptive Statistics >> Crosstabs >> "Statistics"
 Analyze >> Nonparametric Tests >> 2 Related Samples

Yukarıdaki iki seçenekle de McNemar testi yapılabilir. *Nonparametric Tests* seçimi ile *Two-Related Samples Tests* diyalog kutusu açılır (Şekil 8).



Şekil 8
 SPSS'te "Nonparametric tests" seçenekinden "Two-Related Samples Tests" diyalog kutusu.

Soldaki listede dosyadaki sayısal değişkenler yer almaktadır. Değişken çiftlerini seçmek için önce değişkenlerden birini kliklediğinizde, bu değişken adı *Current Selections*'daki *Variable 1*'e aktarılacaktır. İkinci değişkeni kliklediğinizde *Variable 2*'ye bu değişken aktarılacaktır. Daha sonra, bu değişken ölü sağıdaki *Test pairs* listesine aktarılır. *Paired variables* listesine birden çok değişken çifti seçebilirsiniz.

Bu diyalog kutusu ilk açıldığında Wilcoxon testi seçilmiş durumdadır. Bunun yerine McNemar seçeneğini işaretleyerek McNemar testini yapabiliyor siniz.

Örnek 64'teki çalışmanın analizi yapıldığında Şekil 9'daki sonuçlar elde edilecektir.

edilmiştir. Hesaplanan p değeri 0.05'ten küçüktür. Tedavi öncesi ve sonrası AKŞ değerleri arasındaki fark önemlidir.

Test Statistics ^b	
	0. ay AK? & 3. ay AK?
N	100
Chi-Square ^a	16.820
Asymp. Sig.	.000

^a. Continuity Corrected
^b. McNemar Test

Şekil 9 Örnek 64'teki çalışmanın analiz sonucu.

Kappa istatistiği

Analyze >> Descriptive Statistics >> Crosstabs

Crosstabs diyalog kutusunun Statistics altmenüsündeki Kappa seçeneği ile Cohen kappa değeri elde edilir. İki gözlemci arasındaki uyumun testi olduğu için yalnız satır ve sütun değişkenleri aynı değerleri alan kare (2×2 , 3×3 vb.) tablolar için yapılabilir.

Örnek 65'teki çalışmada amaç iki doktor arasında biyopsi preparatlarının değerlendirilmesi açısından fark olup olmadığını incelenmesiydi. Şekil 10'da bu çalışmaya ilgili McNemar testi ve kappa istatistiği görülmektedir. İki doktor arasında fark yoktur ($p=0.58$) ve iki doktor arasındaki uyum iyi derecede ($\kappa=0.615$) ve istatistiksel olarak önemlidir ($p<0.001$).

Dr. Ahmet * Dr. Mehmet Crosstabulation				
Count		Dr. Mehmet		Total
		Benign	Malign	
Dr.	Benign	15	8	23
Ahmet	Malign	5	72	77
Total		20	80	100

Chi-Square Tests		
	Value	Exact Sig. (2-sided)
McNemar Test		.581 ^a
N of Valid Cases	100	

a. Binomial distribution used.

Symmetric Measures				
	Value	Asymp. Std. Err. ^a	Approx. T ^b	Approx. Sig.
Measure of Agreement	Kappa	.615	.096	6.178
N of Valid Cases		100		.000

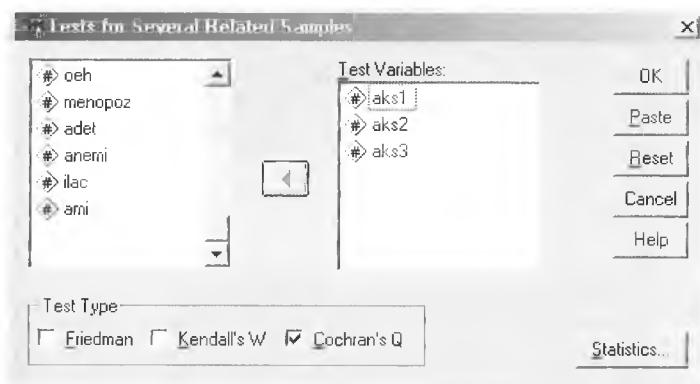
a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

Şekil 10
Örnek 65'teki
çalışmanın
analiz sonucu.

Cochran Q testi

Analyze >> Nonparametric Tests >> K Related Samples

Yukarıdaki seçimle Tests for Several Related Samples diyalog kutusu açılır. Bu diyalog kutusu ilk açıldığında Friedman testi seçilmiş durumdadır. Bu nun yerine Cochran's Q seçeneğini işaretleyerek Cochran Q testini yapabilirsiniz. Bağımlı değişkenleri Test Variables kutusuna seçin ve OK'leyin (Şekil 11).



Şekil 11
SPSS'te
“Nonparametric tests” seçenekinden “Tests for Several Related Samples” diyalog kutusu.

Örnek 64'teki çalışmada tedavinin altıncı ayında da değerlendirme yapıldığını düşünelim. Bu durumda ölçüm sayısı 2'den fazla olduğu için McNemar testi yerine Cochran Q testi yapılmalıdır. Şekil 12'de bu çalışmanın sonuçları görülmektedir. Hesaplanan p değeri 0.05'ten küçüktür. Tedavi öncesi ve 3 ve 6 ay sonrası AKŞ değerleri arasındaki fark önemlidir. İkişerli karşılaştırımlar McNemar testi ile yapılabilir.

Frequencies		
	Value	
	0	1
0. ay AK?	60	40
3. ay AK?	30	70
6. ay AK?	21	79

Test Statistics		
N	100	
Cochran's Q	36.794 ^a	
df	2	
Asymp. Sig.	.000	

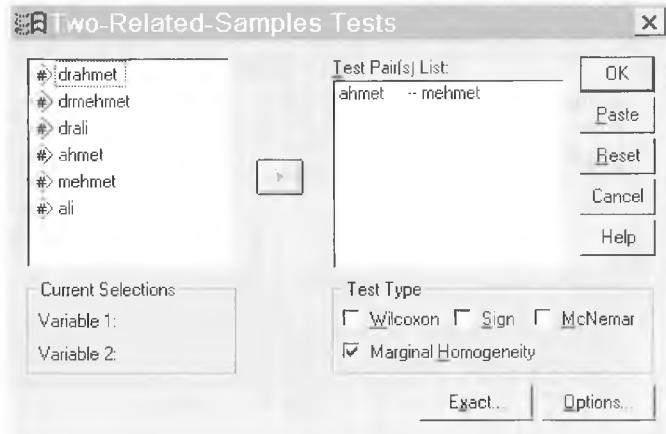
a. 0 is treated as a success.

Şekil 12 Örnek 64'teki çalışmada AKŞ'nın üç kez ölçüldüğü durumda yapılan analiz sonucu.

Marjinal homojenlik testi

Analyze >> Nonparametric Tests >> 2 Related Samples

Yukarıdaki seçimle Two-Related Samples Tests diyalog kutusu açılır. Bu diyalog kutusu ilk açıldığında Wilcoxon testi seçilmiş durumdadır. Bunun yerine Marginal Homogeneity seçeneğini işaretleyerek marjinal homojenlik testini yapabilirsiniz (Şekil 13).



Şekil 13
SPSS'te "Non-parametric tests" seçeneklerinden "Two-Related Samples Tests" diyalog kutusu.

Örnek 66'daki çalışmanın analizi yapıldığında Şekil 14'teki sonuçlar elde edilmiştir. Hesaplanan p değeri 0.05'ten büyüktür ($p=0.38$). İki doktorun splenomegali değerlendirmesi arasındaki fark önemli değildir.

Şekil 14
Örnek 66'daki çalışmanın analiz sonucu.

Marginal Homogeneity Test	
	Dr. Ahmet & Dr. Mehmet
Distinct Values	3
Off-Diagonal Cases	30
Observed MH Statistic	29.000
Mean MH Statistic	33.000
Std. Deviation of MH Statistic	4.583
Std. MH Statistic	-.873
Asymp. Sig. (2-tailed)	.383

BAĞINTI ANALİZİ

bölüm 13

Bağıntı nedir?

Doğrusal olmayan bağıntı

Sayısal değişkenler arasındaki doğrusal bağıntı

Ordinal değişkenler arasındaki doğrusal bağıntı

Bağıntı katsayısının bazı özellikleri

Bağıntı katsayısının önemlilik düzeyi

Bağıntı katsayısı mı, p değeri mi?

Belirtme katsayısı

SPSS'te bağıntı analizi nasıl yapılır?

Daha önceki bölümlerde grup ortalamaları ve oranlarının karşılaştırılmasına ilişkin teknikleri görmüştük. Bu analiz tekniklerinin incelediği asıl konu, grupların ortalamaları ya da oranlarının birbirlerinden farklı olup olmamasıdır. Bölüm 8'de yer alan Örnek 39 ve Bölüm 11'deki Örnek 48'de IL-3 düzeyi ortalamalarının ve ADA düzeyi ortalamalarının değişik gruplar arasında farklılık gösterip göstermediği uygun tekniklerle incelenmiştir.

Peki lösemide IL-3 düzeyi neden normallerden daha yüksektir, lösemideki hangi değişiklik IL-3 düzeyinin yükselmesine neden olmaktadır? Tüberküloza bağlı plevral effüzyonda ADA neden diğer nedenlere bağlı plevral effüzyonlarından daha yüksektir?

Bu soruların yanıtları bağıntı (korelasyon) analizi ve regresyon analizi teknikleri ile araştırılır. Bağıntı analizi, iki sayısal değişken arasında ilişki olup olmadığını araştıran bir tekniktir.

Lösemide IL-3 düzeyinin yüksek olmasına neden olan kanda bulunan blastlar mı, yoksa serum laktik dehidrogenaz (LDH) düzeyinin yüksek olması mıdır, ya da trombosit sayısı ne kadar düşükse IL-3 düzeyi o kadar yüksek midir? Tüm bu spekulatif varsayımların çözümü bağıntı analizi ile başlar, regresyon analizi ile devam eder. Bağıntı analizi ile IL-3 düzeyinin bu üç değişkenle ilişkili olup olmadığı ortaya konulur; regresyon analizi ile de bir hastanın blast sayısı, LDH düzeyi ve trombosit sayısı bilindiğinde, IL-3 düzeyinin ne ölçüde kestirilebileceği test edilir.

Yeri gelmişken, istatistik tekniklerin ne gibi amaçlarla kullanıldığına tekrar değinmeye yarar var. Tanımlayıcı istatistik teknikler, durum saptamasına

yardımcı olur. Analistik istatistik teknikler ise başlıca iki amaçla kullanılır: Birincisi, farkların önemli olup olmadığına incelendiği tekniklerdir. T testleri ve varyans analizi bu grup tekniklerin prototipidir.

Analistik istatistik tekniklerin ikinci amacı ise ilişki saptama ve eldeki verilerden kestirim yapabilme gücünün test edilmesidir. Regresyon analizi ise bu grup tekniklerin prototipidir.

Bölüm 8 Örnek 39'daki çalışmadan bu üç grup analiz teknikleri ile elde edilebilecek sonuçlar Tablo 1'de özetlenmiştir.

Tablo 1 Örnek 39'daki çalışmada kullanılabilecek istatistik analiz teknikleri ve yorumları.

Teknik	Test	Sonuç ve Yorum
Tanımlayıcı istatistik	Ortalama, ortanca, standart sapma vb.	Lösemide ortalama IL-3 düzeyi 12,3 pg/mL'dir.
Farkları inceleyen analistik istatistik teknikleri	Bağımsız iki grup ortalaması için t testi	Lösemide IL-3 düzeyi normalden daha yüksektir.
İlişki ve kestirim inceleyen analistik istatistik teknikleri	Bağıntı analizi ve regresyon analizi	IL-3 düzeyi blast sayısı ile ilişkendir. Blast sayısı arttıkça, IL-3 düzeyi de artmaktadır.

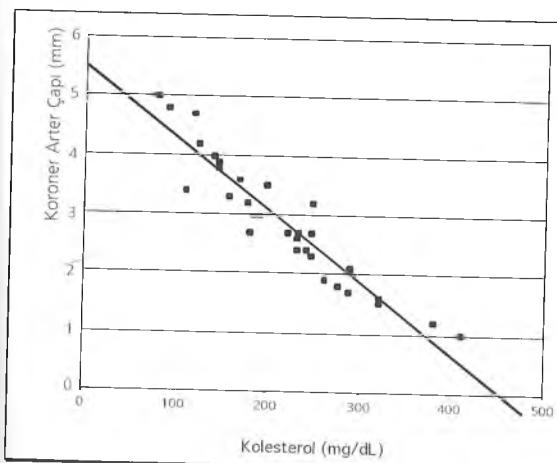
BAĞINTI NEDİR?

ÖRNEK 67

Kolesterol düzeyi ile koroner damarların çapı arasında ilişki olup olmadığı araştırılmak isteniyor. Koroner arter hastalığı tanısı ile koroner anjiyografi yapılan 31 hastada serum kolesterol düzeyi ölçülüyor. Anjiyografi sırasında aterosklerotik plaqın bulunduğu yerde koroner arterin çapı hesaplanıyor.

Acaba kolesterol düzeyi ile koroner arter çapı arasında ilişki var mıdır? Kolesterol düzeyi yüksek olan hastaların koroner arterleri, kolesterol düzeyi normal olanlara göre daha mı dardır? Bu soruyu yanıtlamaya grafik çizerek başlayabiliriz. Yatay ekseni "mg/dL" olarak kolesterol düzeyini, dikey ekseni ise "mm" olarak koroner arter çapını gösterecek şekilde düzenleyelim. Her hastanın yatay eksende kolesterol düzeyine karşılık gelen yerle, dikey eksene

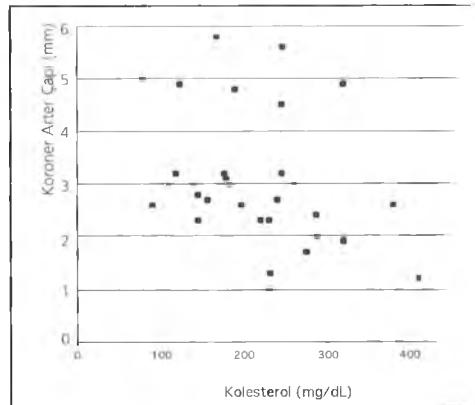
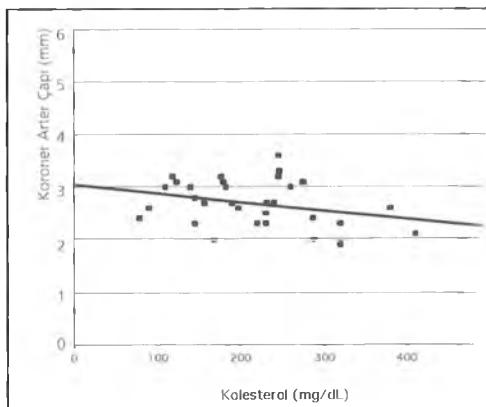
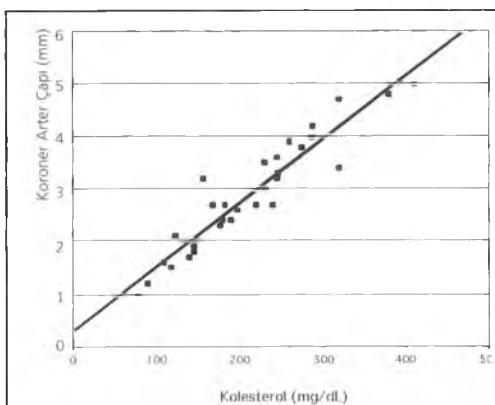
de koroner arter çapına karşılık gelen yerden çizilen hayali dikmelerin kesiştiği yere bir nokta koymalı. Bu tür grafiklere nokta grafiği ya da saçılım grafiği (*scatter-plot*) adı verilir. Şekil 1'de bu çalışmanın sonuçları ile yapılan nokta grafiği gösterilmektedir.



Şekil 1
Örnek 67'deki çalışmada elde edilen kolesterol ve koroner arter çapı ölçümlerinin saçılım grafiği. İki değişken arasında negatif yönde doğrusal bağıntı vardır.

Şekil 1'de görüldüğü gibi kolesterol düzeyi arttıkça koroner arter çapı azalmaktadır. Çünkü hastaları temsil eden noktalar grafiğin sol üst köşesinden sağ alt köşesine doğru uzanan bir doğru etrafında yoğunlaşmaktadır. Bu grafiğe bakarak kolesterol ile koroner arter çapı arasında ters yönde (negatif) bir doğrusal bağıntı olduğu söylenebilir.

Şekil 2'de pozitif doğrusal bağıntı olan ve bağıntı olmayan durumlara örnekler görülmektedir.



Şekil 2

Kolesterol - koroner arter çapına ilişkin saçılım grafikleri. Üstteki grafikte iki değişken arasında pozitif doğrusal bağıntı vardır. Diğer iki grafikte ise iki değişken bağıntılı değildir.

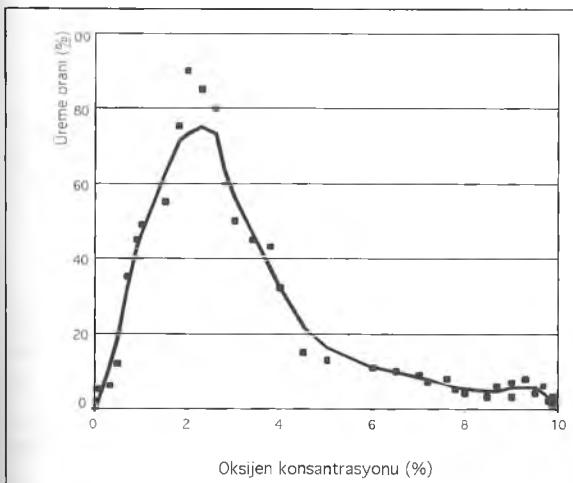
Doğrusal olmayan bağıntı

İki değişken arasında bağıntıdan söz ederken, "doğrusal bağıntı"nın kas-tildiğini vurgulamak gereklidir. Yukarıdaki grafiklerdeki örneklerde noktalar, yatay ya da dikey olmamak koşulu ile bir doğru etrafında toplanıyorlarsa, iki değişken arasında doğrusal bağıntı olduğunu söylemişlik. İki değişken arasındaki bağıntının her zaman doğrusal olması gerekmez.

ÖRNEK 68

Mikroaerofilik mikroorganizmaların üretilmesi sırasında ortamda oksijen konsantrasyonunun % 1-2 civarında tutulması gereklidir. Anaerob ortam ve oksijen konsantrasyonu % 0.5 ile % 10 arasında değişen ortamlarda belirli bir mikroorganizmanın üreme oranı saptanıyor ve Şekil 3'deki grafik çiziliyor. Oksijen konsantrasyonu ile üreme hızı arasında ilişki var mıdır?

Şekil 3'deki noktalar kesinlikle bir doğru etrafında dizilmemiştir. Ancak yine de oksijen konsantrasyonu ile mikroorganizmanın üremesi arasında bağıntı olmadığını söylemek yanlış olur. Çünkü oksijen konsantrasyonun belli bir değerden daha düşük ya da yüksek olması, üremeyi yavaşlatmaktadır, dar bir aralıkta ise optimum mikroorganizma üremesi sağlanmaktadır. Bu durumda iki değişken arasında doğrusal olmayan bağıntı olduğu anlaşılmaktadır.



Şekil 3
Mikroaerofil mikroorganizmaların üreme oranı ile oksijen konsantrasyonu arasındaki bağıntıyı gösteren saçılım grafiği.

Sayısal değişkenler arasındaki doğrusal bağıntı

Değişkenler arasında doğrusal bağıntı olup olmadığını incelemeye geçmeden önce değişkenlerin ölçüm skalalarının belirlenmesi gereklidir. Çünkü nominal ya da ordinal değişkenler arasındaki bağıntının, sayısal değişkenler arasındaki doğrusal bağıntı için kullanılan teknikle araştırılması doğru olmaz.

Yukarıdaki örneklerde kolesterol düzeyi, koroner arter çapı, mikroorganizmaların üreme oranı ve oksijen konsantrasyonu sayısal değişkenlerdir. Sayısal değişkenlerin birbirleriyle ilişkilerinin derecesi Pearson bağıntı (korelasyon) katsayısı (*Pearson coefficient of correlation*) ile belirlenir. Bağıntı katsayısı r ile gösterilir.

Bağıntı katsayısı -1 ile $+1$ arasında değişen değerler alabilir. Nokta grafiği çizildiğinde, denek değerlerine karşılık gelen noktaların bir doğruya ne kadar yakın olup olmadığı bağıntı katsayısı ile belirlenir. Bağıntı katsayısı 0 ise ya noktalar tamamen dağınık durumdadır, ya da eksenlerden birine paralel dizilmişlerdir. Bağıntı katsayısı -1 ya da $+1$ 'e yaklaşıkça noktaların bir doğru üzerinde dizilmiş gibi yerleştiği anlaşıılır. Bağıntı katsayısının başındaki “-” ya da “+” işaretini, bağıntının yönünü belirtir. İki değişkenden birinin değeri arttıkça, diğerinin de arttıyorsa pozitif, azalıyorsa negatif bağıntı var demektir. Örneğin Şekil 1'de bağıntı katsayısı -0.89 , Şekil 2'de birinci grafikte $+0.86$ 'dır. Şekil 2'de ikinci ve üçüncü grafiklerdeki bağıntı katsayısı ise sırasıyla -0.01 ve -0.25 'dir. Bağıntı katsayısı ile ilişkinin derecesi kabaca Tablo 2'de gösterildiği gibidir.

Tablo 2 Bağıntı katsayısının anlamı.

Bağıntı Katsayısı (r)	Anlamı
$0 - 0.25$	Hiç ilişki yok ya da çok zayıf ilişki
$0.25 - 0.50$	Zayıf - orta derecede ilişki
$0.50 - 0.75$	İyi derecede ilişki
$0.75 - 1.00$	Çok iyi derecede ilişki

Biyolojik bilimlerde 0.95 'in üzerinde bağıntı katsayısı bulunmasına çok nadiren rastlanır. Hatta bu durumda bir hata yapılmış yapılmadığı gözden geçirilmelidir. Eğer iki değişken arasında gözden kaçan bir bağımlılık varsa bu tip bir artefakt olabilir. Örneğin aynı vücut ağırlığına sahip kişilerin boy ve vücut kitlesi indeksleri arasındaki bağıntı katsayısı 1.00 'e yakın olacaktır. (Vü-

cut kitlesi indeksi = Vücut ağırlığı / Boy). Bunun nedeni vücut kitlesi indeksinin hesaplanma formülünde boy uzunluğunun da bulunmasıdır. Yani iki değişken birbirine bağlımlıdır.

Ordinal değişkenler arasındaki doğrusal bağıntı

Ordinal değişkenler arasındaki bağıntının incelenmesinde kullanılan teknik Spearman bağıntı analizi'dir. Dağılımı basık ya da üç değerleri bulunan sayısal değişkenler için de bu teknik daha uygundur.

Spearman bağıntı katsayısı r_s olarak kısaltılır. Spearman bağıntı katsayısı da Pearson bağıntı katsayısı gibi -1 ile +1 arasında değerler alır. Yorumlanması da aynıdır. Örneğin Apgar skoru ile doğum ağırlığı arasındaki ilişkinin derecesini belirlemek için Spearman bağıntı katsayısı kullanılmalıdır. Çünkü Apgar skoru ordinal skala'da ölçülen bir değişkendir.

Aralarındaki bağıntının araştırıldığı iki değişkenden birinin ordinal, ya da simetrik olmayan dağılım gösteren sayısal değişken olması durumunda, diğer değişkenin ölçüm skala'sı ne olursa olsun, Spearman tekniği ile bağıntı analizi yapılmalıdır.

Bağıntı katsayısının bazı özellikleri

Bağıntı katsayısı, değişkenlerin ölçüldüğü birimlerden etkilenmez. Örneğin yaşla Hb düzeyi arasında bağıntı arandığında, yaşı yıl, ay, hafta ya da gün olarak hesaplamak r değerini değiştirmez. Ya da koroner arter çapının mm yerine cm olarak verilmesi, bağıntı katsayısını değiştirmez.

Bağıntı katsayısı, simetrik bir ölçütür. Yani nokta grafikte hangi değişkenin yatay, hangi değişkenin dikey eksende olduğu önemli değildir.

Bağıntı katsayısı da, standart sapma gibi uc değerlerden çok etkilenir. Örneğin kolesterol-damar çapı ilişkisinde gruba kolesterol ya da damar çapı grubun değerlerinden çok büyük ya da küçük olan bir denek eklense, bağıntı katsayısı önemli derecede etkilenir. Bu nedenle değişkenlerin değerlerinin dağılımı basıksa, ya da üç değerler varsa, bağıntı katsayısı iki değişken arasındaki ilişkiyi iyi temsil etmez. Bu durumda değişkenlere dönüşüm uygulamak gereklidir.

Bağıntı katsayısı yalnızca doğrusal ilişkinin derecesini belirtir. İki değişken arasında doğrusal olmayan, ama kuvvetli bir ilişki olduğu halde, bağıntı katsayısı düşük olabilir. Örneğin ayın günleriyle hastaneye yatan hasta sayısı arasında bağıntı katsayısı küçük olduğu halde, iki değişken arasında doğrusal olmayan bir ilişki olduğu kesindir. Her yedi günün iki gününde (tatil günlerinde) yatan hasta sayısı diğer günlere göre az olacaktır. Bu nedenle bağıntı araştırılırken, mutlaka nokta grafik de çizilmelidir.

Son olarak, bağıntı katsayısının büyük bulunması, tek basına neden-sonuc ilişkisini açıklamaz. Altta yatan mantıksal bir temeli olmadan, büyük bulunan bir bağıntı katsayısı ile bir değişkenin diğer değişkeni etkilediği ya da ondaki değişimin nedeni olduğu yorumuna gidilmemelidir.

Bağıntı katsayısının önemlilik düzeyi

Bağıntı katsayısı +1 ya da -1'e ne kadar yakınsa, iki değişken arasında o kadar yakın ilişki ya da bağıntı var demektir. Tersine 0'a yakın bir bağıntı katsayısı, iki değişken arasında bağıntı olmadığı ya da çok zayıf bağıntı olduğu anlamına gelir. Başka bir deyişle bağıntı katsayısının 0'dan önemli derecede farklı olması, bağıntının olduğu anlamına gelir. Peki 50 denekli bir çalışmada bağıntı katsayısı 0.09 olarak bulunursa, bağıntı önemli midir? Ya da 12 denekli bir çalışmada 0.34 olarak bulunan bir bağıntı katsayısı, iki değişken arasında önemli bir bağıntı bulunduğu anlamına gelir mi?

Her ne kadar Tablo 2'deki gibi yorumlama bağıntının derecesi hakkında kabaca bir fikir verse de, bağıntı katsayısının 0'dan farklı olduğunu test edilmesi gereklidir. Bu nedenle bağıntı analizi yapıldığı zaman, bağıntı katsayıları ile birlikte, bu katsayıların önemlilik düzeyini belirten p değerleri de bilinmelidir. 0.05'ten daha küçük p değerleri, bağıntı katsayısının 0'dan önemli derecede farklı olduğunu, yani iki değişken arasındaki bağıntının istatistiksel olarak önemli olduğunu gösterir.

Bağıntı katsayısının p değeri, denek sayısı ile doğrudan ilişkilidir. Denek sayısı arttıkça, p değeri küçülür, yani bağıntının önemlilik derecesi artar. Örneğin 67'deki çalışmanın 10 hasta ile yapıldığını ve serum kolesterol düzeyi ile koroner arter çapı arasındaki bağıntı katsayısının -0.47 olarak bulunduğu varsayıyalım. Tablo 2'deki yorumla göre iki değişken arasında zayıf-iyi arasında bağıntı var demektir. Ancak p değeri 0.17 olarak bulunmuştur; yani bağıntı katsayısı 0'dan farklı değildir. Aynı çalışmayı 20 ve 40 kişi ile tekrarlayıp, benzer sonuçların elde edildiğini düşünelim. Bu kez bağıntı katsayısı yine 0.47 olduğu halde, p değeri, sırasıyla 0.04 ve 0.002 olarak bulunacaktır. Göründüğü gibi çalışma 20 hasta ile yapılrsa, aynı büyüklükteki bağıntı katsayısı önemli derecede 0'dan farklı olurken, hasta sayısı 40'a çıkarılırsa, bağıntı katsayısının p değeri çok küçülmektedir.

Bağıntı katsayısının p değerinin denek sayısı arttıkça giderek küçülmesi nedeniyle, özellikle denek sayısı çok fazla olan çalışmalarda küçük bağıntı katsayılarının önemli olduğu şeklinde p değerleri elde edilir. Tersine denek sayısı az olan çalışmalarda p değeri önemsiz (>0.05) bile olsa, büyük bağıntı katsayıları iki değişken arasında bağıntı olduğunu işaret edebilir.

ÖRNEK 69

Yaşla eritrosit sedimentasyon hızının (ESH) ilişkili olup olmadığını araştırmak amacıyla çeşitli yaşlardan 600 sağlıklı kişi seçiliyor ve ESH ölçülüyor. Yaşla ESH arasındaki bağıntı katsayısı (r) 0.12 ve bağıntı katsayısının p değeri 0.005 bulunuyor.

ÖRNEK 70

Toplumların gelir düzeyi ile sağlık düzeyi arasında bağıntı olup olmadığını araştırmak amacıyla seçilen 12 köyde kişi başına düşen ortalama gelir ile bebek ölüm hızları hesaplanıyor. İki değişken arasındaki bağıntı katsayısı (r) -0.48 ve bağıntı katsayısının p değeri 0.12 bulunuyor.

Örnek 69'daki çalışmada bağıntı katsayısının p değeri <0.05 olduğuna göre yaş arttıkça ESH da artmaktadır ve iki değişken arasındaki bağıntı önemli gibi görülmektedir. Ancak bağıntı katsayısı 0.12, çok küçük olduğuna göre bu sonuç aslında hasta sayısının fazla olmasına bağlı bir artefakt olarak yorumlanabilir.

Örnek 70'deki çalışmada ise denek sayısı (her köy bir denektir) az olduğu için, -0.48 gibi, 0'dan oldukça farklı bir bağıntı katsayısına karşın, p değeri >0.05 olarak bulunmaktadır. Buradaki sonuç ise denek sayısının az olmasına bağlı bir artefakttır. Eğer aynı özellikte başka 12 köy de çalışmaya alınsaydı, p değeri 0.02 olarak bulunacak ve gelir düzeyi arttıkça, bebek ölüm hızının azalduğu sonucuna ulaşılacaktı.

Bağıntı katsayısı mı, p değeri mi?

İstatistikte çoğu kez p değeri diğer her türlü parametreden önde gelirken, bağıntı analizinde bağıntı katsayısının değeri, p değerinden daha çok bilgi verir. Bu nedenle bağıntı katsayısı ve p değerlerini incelerken, denek sayılarına dikkat etmek gereklidir.

Az sayıda denek içeren bir çalışmada yeterince büyük bağıntı katsayısına karşın $p>0.05$ ise, hemen bağıntının önemsiz olduğu yorumuna gitmemek gereklidir. Aynı şekilde fazla sayıda denekle yapılan çalışmalarında p değerinden çok bağıntı katsayısının değerine dikkat etmek daha doğru olur.

Bu ilkeleri literatürdeki çalışmaları izlerken göz önüne almak gereklidir. Öte yandan yaptığınız çalışmalarında da denek sayısı yetersiz ya da çok fazla ise, yalnız p değerine dikkat ederek hatalı yorumlara gidebilirsiniz.

Örneğin 10 denekli bir çalışmada $r=0.58$ olmasına karşın, p değerinin

0.05'den büyük olarak bulunması arasındaki çelişki, denek sayısının arttırılması ile çözülebilir. Öte yandan 300 denekli bir çalışmada $r=0.09$ ve $p=0.11$ bulunduğunda, denek sayısı daha da arttırılınca p değerinin küçüleceği kesin olmakla birlikte, bunun bir artefakt olduğunun bilinmesi ile gereksiz yere emek ve para harcanmasından kaçınılmış olur.

Belirtme katsayısı

Bağıntı katsayısının karesi (r^2), belirtme katsayısı (*coefficient of determination*) olarak adlandırılır. Örneğin Şekil 1'deki değerler için kolesterolle damar çapı arasındaki bağıntı katsayısı $r=-0.89$ olduğuna göre, belirtme katsayısı, $r^2=0.79$ 'dur. Bağıntı katsayısı -1 'e yakın olduğuna göre kolesterol değeri arttıkça, damar çapı azalmaktadır.

Belirtme katsayısı, değişkenlerden birinin değeri bilindiğinde diğer değişkenin değişkenliğinin, yani varyansının ne kadar azaldığını gösterir. Yani hastaların kolesterol değeri bilindiğinde, damar çapının değişkenliği % 79 azalmaktadır. Bu ne anlama gelir?

Şekil 1'den Örnek 67'deki çalışmaya alınan hastaların koroner arter çaplarının 1 ile 6 mm arasında değiştiği ve kolesterolle negatif bağıntılı olduğu anlaşılmaktadır. Ortalama koroner arter çapı 3.5 mm ve standart sapma 1.5 mm ise (bu durumda varyans $1.52 = 2.25$ olacaktır), hastaların % 95'inin koroner arter çapının 0.5 ile 6.5 mm arasında olduğu varsayılabılır (Normal dağılım olduğu takdirde hasta değerlerinin % 95'inin ortalaması $\pm 2SD$ içinde olduğu nu hatırlayınız). Kolesterol düzeyi 200 mg / dL olan bir hastanın koroner arter çapını kestirmeye çalışırsak, % 95 olasılıkla 0.5 ile 6.5 mm olabileceğini öne sürebiliriz. Kolesterol ile koroner arter çapı arasındaki bağıntının belirtme katsayısı biliniyorsa, bu aralık biraz daha daraltılabilir.

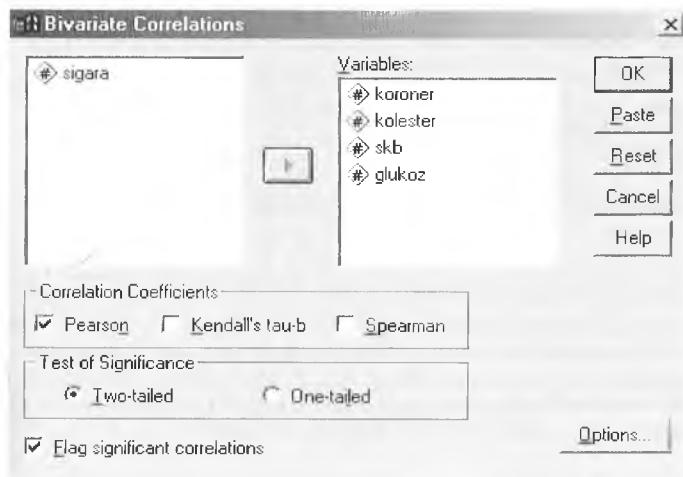
İşte belirtme katsayısı bu aralığın ne kadar daraltılabilenin bir ölçüsüdür. Şekil 1'de 200 mg / dL kolesterol düzeyine karşılık gelen koroner arter çapı 3 mm'dir. Tüm grubun varyansı 2.25 ve belirtme katsayısı 0.79 olduğuna göre, kolesterol düzeyi bilinen bir hastanın olası koroner arter çapının yayılma ölçütü olarak varyans % 79 azalmış ve 0.47 olmuştur. Standart sapma ise 1.5 mm'den 0.7 mm'ye azalmıştır. O kalde kolesterol düzeyi 200 mg / dL olan bu hastanın koroner arter çapı, % 95 olasılıkla 3 mm'den iki standart sapma kadar daha dar ya da geniş olabilir, yani 1.6 ile 4.4 mm arasında olmalıdır.

SPSS'TE BAĞINTI ANALİZİ NASIL YAPILIR?

Analyze >> Correlate >> Bivariate

Yukarıdaki seçimle **Bivariate Correlations** diyalog kutusu açılır (Şekil 1). Soldaki listede yalnız sayısal değişkenler vardır. *Variables* kutusuna bağıntı ana

lizine alınacak en az iki değişken seçilir. *Correlation Coefficients* kutusundaki Pearson, Kendall tau-b ve Spearman bağıntı katsayıları seçeneklerinden en az birinin seçilmiş olması koşulu ile OK'lerseniz, bağıntı analizi yapılacaktır.



Şekil 4
SPSS'te "Bivariate Correlations" diyalog kutusu.

Pearson bağıntı analizinin uygulama koşulları sağlanamıyorsa, Kendall tau-b ya da Spearman bağıntı katsayıları seçilmelidir. Kendall tau-b de, Spearman bağıntı katsayısı gibi nonparametrik bağıntı analizi ile elde edilen bir bağıntı katsayısıdır.

Test of significance kutusundan tek-yönlü ya da iki-yönlü hipotez testi seçeneklerinden birini seçebilirsiniz. Eğer önceden bağıntı katsayısının yönü hakkında bir öngörünüz yoksa, yani bağıntı katsayısının pozitif mi, negatif mi olduğu hakkında bir kestirim yapamıyorsanız, iki-yönlü hipotez testinin kullanılması doğru olacaktır. Zaten bu seçenek önceden seçilmişdir. *Flag significant correlations* seçeneği de önceden seçilmiştir. Bu durumda bağıntı tablosunda bağıntı katsayılarının 0'a eşit olma olasılığı anlamına gelen p değerleri 0.05'ten küçükse tek asteriksle (*), 0.01'den küçükse çift asteriksle (**) işaretlenir.

Şekil 5'de, Örnek 67'deki çalışmada değişkenlere ilişkin Pearson bağıntı katsayıları görülmektedir. Her sıradaki ilk rakamlar r değerlerine, ikinci rakamlar p değerlerine ve en alttaki sayılar ise hasta sayılara karşılık gelmektedir. Örneğin sistolik kan basıncı ile koroner arter çapı arasındaki bağıntı katsayısı -0.358 ve p değeri 0.048'dir. Öte yandan glukoz ile koroner arter çapı arasındaki bağıntı katsayısı 0'a yakın (-0.162) ve dolayısıyla karşılık gelen p değeri de 0.05'ten büyuktur (0.384).

Correlations					
		KORONER	KOLESTER	SKB	GLUKOZ
KORONER	Pearson Correlation	1.000	-.941	-.358	-.162
	Sig. (2-tailed)		.000	.048	.384
	N	31	31	31	31
KOLESTER	Pearson Correlation	-.941	1.000	.323	.219
	Sig. (2-tailed)	.000		.077	.237
	N	31	31	31	31
SKB	Pearson Correlation	-.358	.323	.219	.261
	Sig. (2-tailed)	.048	.077	.237	.156
	N	31	31	31	31
GLUKOZ	Pearson Correlation	-.162	.219	.261	1.000
	Sig. (2-tailed)	.384	.237	.156	
	N	31	31	31	31

Şekil 5 Örnek 67'deki çalışmada koroner arter çapı, kolesterol düzeyi, sistolik kan basıncı ve glukoz düzeyi arasında ikili bağıntı analizi ile elde edilen Pearson bağıntı katsayıları. Tabloda satır ve sütunların kesiştiği yerdeki üç rakam sırasıyla bağıntı katsayısı, p değeri ve hasta sayısını vermektedir. Örneğin sistolik kan basıncı ile koroner arter çapı arasındaki bağıntı için Pearson bağıntı kat sayısının -0.358, hasta sayısının 31 ve p değerinin 0.048 olduğu görülmektedir.

Şimdi yukarıdaki değişkenlerin normal dağıldığını ya da üç değerlerin bulunduğu varsayıyalım. Bu durumda Pearson yerine Spearman bağıntı kat sayısının kullanılması gereklidir. Sonuçlar Şekil 6'da gösterilmiştir.

Correlations					
		KORONER	KOLESTER	SKR	GLUKOZ
Spearman's rho	KORONER	Correlation Coefficient	1.000	-.939	-.302
		Sig. (2-tailed)	.000	.098	.630
		N	31	31	31
	KOLESTER	Correlation Coefficient	-.939	1.000	.313
		Sig. (2-tailed)	.000	.086	.720
		N	31	31	31
	SKB	Correlation Coefficient	-.302	.313	1.000
		Sig. (2-tailed)	.098	.086	.154
		N	31	31	31
	GLUKOZ	Correlation Coefficient	-.092	.066	.262
		Sig. (2-tailed)	.630	.720	.154
		N	31	31	31

Şekil 6 Örnek 67'deki çalışmada koroner arter çapı, kolesterol düzeyi, sistolik kan basıncı ve glukoz düzeyi arasında ikili bağıntı analizi ile elde edilen Spearman bağıntı katsayıları. Tabloda satır ve sütunların kesiştiği yerdeki üç rakam sırasıyla bağıntı katsayısı, p değeri ve hasta sayısını vermektedir. Örneğin sistolik kan basıncı ile koroner arter çapı arasındaki bağıntı için Spearman bağıntı kat sayısının -0.302, hasta sayısının 31 ve p değerinin 0.098 olduğu görülmektedir.

Kolesterol ve glukoz ile koroner arter çapı arasındaki bağıntı katsayılarına karşılık gelen p değerlerinde bir miktar değişiklik olsa da sonuçta genel olarak önemli bir değişiklik olmamıştır (Bunun nedeni büyük olasılıkla bu değişkenlerin dağılıminin normale yakın olması ve ayrıca parametrik bağıntı analizinin koşullarının sağlanıyor olmasıdır).

Öte yandan sistolik kan basıncı ile koroner arter çapı arasındaki bağıntı katsayısına karşılık gelen p değeri parametrik bağıntı analizi ile 0.05'ten küçükken nonparametrik bağıntı analizi ile 0.05'ten büyük bulunmuştur. Bu durumda bu iki değişken arasındaki bağıntının parametrik ya da nonparametrik yöntemlerden hangisi ile incelenmesi gerektiğine karar vermek için test varsayımlarının gözden geçirilmesi gerekecektir.

bölüm 14

REGRESYON ANALİZİ

Regresyon katsayısının bazı özellikleri

Regresyon analizinin varsayımları

Varsayımların sağlanıp sağlanmadığı nasıl kontrol edilir?

Rezidüel analizi

Bağımsız değişken sayısı birden fazla ise

Regresyon eşitliğine alınacak bağımsız değişkenleri nasıl seçelim?

Forward selection

Backward elimination

Stepwise selection

SPSS'te regresyon analizi nasıl yapılır?

İki değişken arasındaki bağıntıyı incelerken nokta grafik çizmenin yararlı, hatta gerekli olduğundan söz etmişlik. Bölüm 13'te verilen Örnek 67'deki çalışmada verilere dayanarak çizilen Bölüm 13'te yer alan Şekil 1'deki grafiğe dönemlim.

Bu grafikteki eğri sol üstten sağ alta doğru uzandığına göre, kolesterol ile koroner arter çapı arasında negatif bir bağıntı olduğunu artık biliyoruz. Grafikteki doğru dikey ekseni koroner arter çapının 6 mm olduğu yerden kesmektedir. Yatay eksen iseコレsterol düzeyinin 400 mg/dL olduğu yerden kesilmektedir; yanıコレsterol düzeyindeki her 100 mg/dL artış, koroner arter düzeyinde 1.5 mm daralmaya karşılık gelmektedir.

Bu grafiğe bakarakコレsterol düzeyi 0 mg/dL olabileceydi, koroner arter çapının 6 mm olacağı varsayılabildi. Öte yandanコレsterol düzeyinin 400 mg/dL'ye eşit ya da daha fazla olması durumunda ise koroner arter çapının 0 mm olacağı öne sürülebilirdi. Tabii kiコレsterol düzeyi 0 mg/dL olamaz, ama bu üç noktaları bir tarafa bırakırsak, Bölüm 13 Şekil 1'deki doğru,コレsterol ile koroner arter çapı arasındaki ilişkiyi iyi bir şekilde ortaya koymaktadır.

Pekiコレsterol düzeyi 250 mg/dL olduğu zaman koroner arter çapının kaç mm olması beklenir, ya da koroner arter çapı 3.4 mm olduğundaコレsterol düzeyi kaç olarak kestirilebilir? Bu sorulara grafikten aşağı yukarı doğru yanıtlar bulunabilir. Kesin yanıt ise aşağıdaki formülden elde edilebilir:

$$\text{Koroner arter çapı (mm)} = 6 - (0.015 \times \text{Kolesterol (mg/dL)})$$

Bu formül nasıl elde edildi? Bu formüldeki "6", grafikteki doğrunun dikey ekseni kestiği yer. "0.015" ise $6 / 400$, yani doğrunun eğimi. Doğrunun eğimi negatif olduğu için 0.015 'in önünde " $-$ " işaretü var.

Bir önceki paragraftaki sayılarla deneyelim. Kolesterol düzeyi 0 mg / dL olduğunda, koroner arter çapı $= 6 - (0.015 \times 0) = 6 \text{ mm}$ oluyor. Kolesterol düzeyi 400 mg / dL olduğunda ise, koroner arter çapı $= 6 - (0.015 \times 400) = 6 - 6 = 0 \text{ mm}$ oluyor. Artık kolesterol düzeyinin her değerine karşılık gelmesi beklenen koroner arter çapını bu formülden hesaplayabiliriz.

Regresyon analizi, sayısal değişkenler arasındaki ilişkinin derecesini ve
ren bağımlı analizine ek olarak, değişkenlerden biri bilindiğinde, diğerinin
değerini kestirebilmemize yarayan bir istatistik analiz tekniğidir. Kestirim
yapabilmemizi sağlayan formüle ya da eşitliğe regresyon eşitliği adı verilir.
Regresyon eşitliğinin formüle ettiği doğruya ise regresyon doğrusu adı veri-
tir. Regresyon eşitliği şu şekildedir.

$$Y = a + bX \quad \text{ya da} \quad Y = \beta_0 + \beta_1 X$$

Bu eşitlikte "Y", bağımlı değişken, yani hakkında kestirim yapılacak değişkendir. Yukarıdaki örnek için "koroner arter çapı" bağımlı değişkendir.

X ise bağımsız değişkendir, yani Y'nin değerinin kestiriminde yararlanılacak değişkendir.

"a" dikey ekseni kesim noktasıdır. Buna sabit (*constant*) ya da kesim noktası (*intercept*) adı verilir ve β_0 ile gösterilir.

"b" ise regresyon doğrusunun eğimidir. Buna regresyon katsayı adı verilir ve β_1 ile gösterilir. Regresyon analizinde bu iki parametrenin önemlilik testi yapılır. Birincisi β_0 'ın 0 'dan farklı olup olmadığı, yani regresyon doğrusunun dikey ekseni kestiği yerin 0 olup olmadığı test edilir.

Eğer p değeri <0.05 ise, a'nın $\%95$ olasılıkla 0 olmadığı sonucuna varılır. Aslında aradığımız ya da ilgilendirdiğimiz genellikle bu değildir. İki değişken arasında ilişki aradığımıza göre, asıl ilgilendirdiğimiz β_1 'in değeridir. Çünkü eğer β_1 , 0 'a çok yakın bir değerse, formüle dikkat ederseniz, değişkenlerden biri, diğerinin kestirilmesinde yararlı olmayacağındır, başka bir deyişle iki değişken bağıntılı değildir. Bu nedenle β_1 'in 0 'dan farklı olup olmadığı önemlidir. İşte regresyon analizinde test edilen diğer hipotez ise β_1 'in 0 olup olmadığıdır. Eğer p <0.05 ise b değeri 0 'dan farklıdır, yani iki değişkenin arasındaki bağıntı önemlidir.

		KORONER			
KOLESTER	Pearson Correlation				
	Sig. (2-tailed)				
N				31	
Coefficients ^a					
Model	Unstandardized Coefficients		Standardi zed Coefficien ts	t	Sig
	B	Std. Error			
1 (Constant)	5.506	.185		29.756	.000
KOLESTER	-1.22E-02	.001	-.941	-15.029	.000

^a Dependent Variable: KORONER

Şekil 1 Kolesterol - koroner arter çapı arasındaki ilişkinin regresyon analizi.

Şekil 1'de kolesterol ile koroner arter çapı arasındaki bağıntı katsayısı ve regresyon eşitliğine ait önemlilik testlerinin sonuçları görülmektedir.

Gördüğü gibi bağıntı katsayısı -0.941'dir, yani çok iyi derecede negatif bağıntı vardır. Regresyon tablosunda kolesterol yazan satırda "B" ile gösterilen -1.22E-02 (yani -0.0122) kolesterol değişkeninin regresyon katsayısıdır.

Aynı satırın en sonunda Sig ile gösterilen değer ise regresyon katsayısının 0'a eşit olup olmadığı belirten p değeridir. P değeri <0.0001 olduğuna göre kolesterol değişkeninin regresyon katsayısı olan -0.0122'nin 0'dan farklı olduğu anlaşılmaktadır; yani kolesterol düzeyi ile koroner arter çapı arasında doğrusal ilişki vardır ve kolesterol düzeyinden koroner arter çapı kesitilebilir.

Constant yazan satırda B'ye karşılık gelen 5.51, regresyon doğrusunun dikay eksenini, yani koroner arter çapını gösteren eksenin nereden kestiğini göstermektedir. Bu satırın sonundaki p değeri de <0.0001 olduğu için bu noktanın 0'a eşit olmadığı anlaşılmaktadır.

Bu tablodaki değerlerden regresyon eşitliği şöyle yazılabilir:

$$\text{Koroner arter çapı (mm)} = 5.51 - (0.0122 \times \text{Kolesterol (mg/dL)})$$

REGRESYON KATSAYISININ BAZI ÖZELLİKLERİ

Regresyon katsayısının önemi ile bağıntı katsayısının önemi paraleldir. Yani iki değişken arasındaki bağıntı, bağıntı analizi ile önemli bulunurken, regresyon analizi ile önemsiز bulunamaz. Çünkü regresyon analizi, iki değişken arasındaki bağıntının derecesini irdelerken, aslında bağıntı analizine dayanmaktadır.

Regresyon katsayısi, değişkenlerin ölçüldüğü birimlerden etkilenir. Değişkenlerin birimleri değişikçe, regresyon analizindeki katsayılar da değişir. Yukarıdaki tablodaki regresyon katsayısi ve kesim noktası değeri, kolesterol ve koroner arter çapının ölçüm birimlerine doğrudan bağlıdır. Örneğin koroner arter çapı mm olarak değil de, cm olarak ölçülmüş olsaydı, kolesterol değişkeninin regresyon katsayısi - 0.00122 ve kesim noktası 0.551 olacaktır. Ancak hesaplanan p değeri değişkenlerin birimlerinden etkilenmez. Regresyon katsayıları değişkenlerin birimlerine göre değiştiği için, regresyon katsayısının büyülüğüne bakıp, iki değişken arasındaki ilişkinin derecesini kestirmek olası değildir. Yukarıdaki örnekteki regresyon katsayıları (-0.0122 ve -0.00122) 0'a yakın gibi görülmekte birlikte, regresyon analizi sonucunda 0'dan önemli derecede farklı olduğu anlaşılmıştır. Tersine regresyon katsayısi büyük olduğu halde, karşılık gelen p değeri >0.05 olabilir, yani değişkenler arasındaki ilişki ötemsiz olabilir.

Regresyon katsayısi, simetrik bir ölçüt değildir. İki değişkenden hangisinin bağımsız, hangisinin bağımlı olduğuna göre regresyon katsayısi ve kesim noktası değeri değişir. Örneğin kolesterol düzeyinden (bağımsız değişken) koroner arter capını (bağımlı değişken) kestirmek yerine, koroner arter capından kolesterol düzeyini kestirmek istiyorsak hesaplanan regresyon katsayısi ve kesim noktası değişecek ve regresyon eşitliği aşağıdaki gibi olacaktır. Ancak ilki değişken arasındaki ilişkinin önemini belirten p değeri değişimeyecektir.

$$\text{Kolesterol düzeyi (mg/dL)} = 425 - (72.77 \times \text{Koroner arter çapı (mm)})$$

Regresyon katsayısı da, bağıntı katsayısı gibi uç değerlerden çok etkilenir. Örneğin kolesterol-damar çapı ilişkisinde gruba kolesterol ya da damar çapı grubun değerlerinden çok büyük ya da küçük olan bir hasta eklense, regresyon katsayısı önemli derecede etkilenir. Bu nedenle değişkenlerin değerlerinin dağılımı basıkça, ya da uç değerler varsa, regresyon analizi yapılamaz. Bu durumda değişkenlere dönüşüm uygulamak gereklidir.

Regresyon katsayısi yalnızca doğrusal ilişkinin derecesini belirtir (?). Regresyon eşitliği $y = a + bX$ şeklinde ise bu cümle doğrudur. Ancak doğrusal ol-

mayan ilişkileri de analiz etmek olasıdır. Bu nedenle regresyon eşitliği (ya da regresyon modeli) doğrusal olmayan ilişkileri analiz etmek amacıyla da kullanılabılır.

Son olarak, regresyon katsayısının önemli bulunması, tek başına neden-sonuc ilişkisini açıklamaz. Altta yatan mantıksal bir temeli olmadan, önemli bulunan bir regresyon katsayı ile bir değişkenin diğer değişkeni etkilediği ya da ondaki değişimin nedeni olduğu yorumuna gidilmemelidir.

REGRESYON ANALİZİNİN VARSAYIMLARI

Bağıntı analizinde değişkenlerin normal dağılmadığı ya da uç değerlerin bulunduğu durumlarda ya da ölçüm düzeylerinin ordinal olduğu durumlarda Pearson bağıntı analizi yerine nonparametrik karşılığı olan Spearman bağıntı analizinin yapıldığını görmüştük. Regresyon analizinin uygulanabilmesi için de benzer varsayımlar vardır. Öncelikle iki değişkenin dağılımının normal olması gereklidir. Bunun dışında iki değişkenin ikili normal dağılıma (*bivariate normal distribution*) uyması gereklidir. Yani X değişkeninin her değerine karşılık gelen Y değerlerinin normal dağılması gereklidir. Örneğin kolesterol değeri 100-125 mg/dL arasında olan hastaların koroner arter çaplarının normal dağılması, 125-150 mg/dL arasında olan hastaların koroner arter çaplarının yine normal dağılması, ... gereklidir.

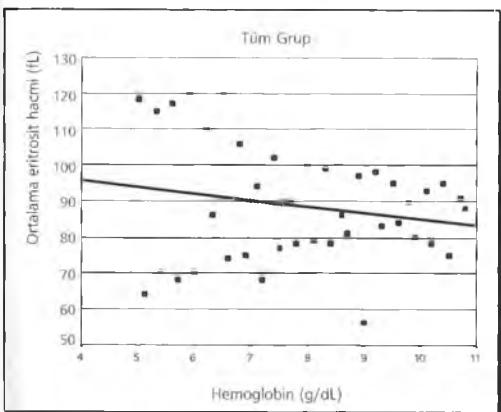
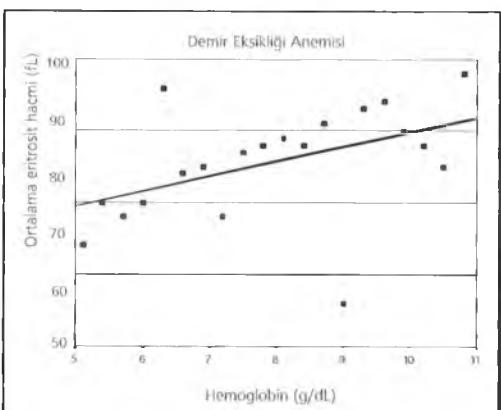
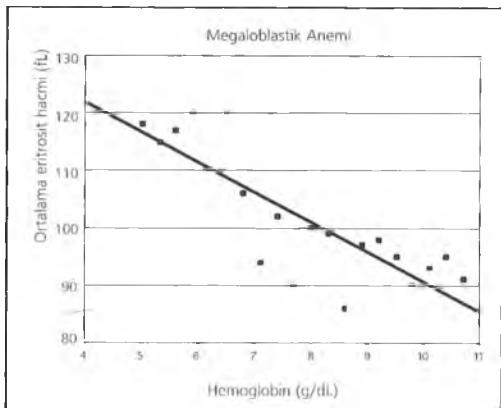
Bir diğer varsayımda X değişkeninin her değerine karşılık gelen Y değerlerinin standart sapmalarının aynı olmasıdır. Örneğin kolesterol değeri 100-125 mg/dL arasında olan hastaların koroner arter çaplarının standart sapması ile, 125-150 mg/dL arasında olan hastaların koroner arter çaplarının standart sapmasının, ... yaklaşık aynı olması gereklidir. Bu varsayıma homojenlik varsayıımı denir. Varyans analizindeki varyansların homojenliği varsayıımına benzettilebilir.

Regresyon analizinin varsayımlarından biri de iki değişken arasındaki ilişkinin doğrusal olmasıdır. Bu varsayımda sağlanamadığı takdirde ya değişkenlere dönüşüm uygulayarak, ilişki doğrusallaştırılmalıdır, ya da regresyon modelinde değişiklik yaparak doğrusal olmayan ilişki incelemelidir.

Regresyon analizi ile incelenenek veriler tek bir gruba ait olmalıdır. Birden çok grup birleştirildiği takdirde, gerçekte var olan ilişki gruplarının birleşmesi sonucunda gözden kaçabilir ya da gerçekte aralarında ilişki bulunmayan değişkenler arasında artefakt olarak ilişki saptanabilir.

Hb düzeyi ile ortalama eritrosit hacmi (OEH) arasında ilişki olup olmadığı araştırılıyor. Çalışmaya megaloblastik anemi ve demir eksikliği anemisi olan hastalar alınmış. Megaloblastik anemide Hb değeri azaldıkça, OEH'nin arttığı ($r = -0.83$), demir eksikliği anemisinde ise Hb değeri azaldıkça, OEH'nin de azaldığı ($r = 0.46$) görülüyor. Tüm hastalar dahil edilerek yapılan regresyon analizinde ise Hb ile OEH arasında zayıf ilişki olduğu saptanıyor ($r = -0.19$).

Örnek 71, farklı özelliklerdeki grupların birleştirilmesi ile gerçekte var olan bir ilişkinin gözden kaçabileceğini gösteren iyi bir örnektir. Değişkenler arasındaki ilişki farklı gruplarda farklı yönde olduğu zaman, grupların birleştirilmesi ile ilişkilerin yönleri birbirlerini nötralize ederler ve değişkenler arasında ilişki olmadığı gibi hatalı bir sonuç elde edilebilir. Şekil 2'de her iki gruba ait regresyon doğrularının farklı yönde olduğu ve iki grubun birleştirildiği durumda elde edilen regresyon doğrusunun ise yatay eksene paralel olduğu, yani OEH ile Hb arasında ilişki olmadığı ya da zayıf ilişki olduğu anlamına geldiği görülmektedir.

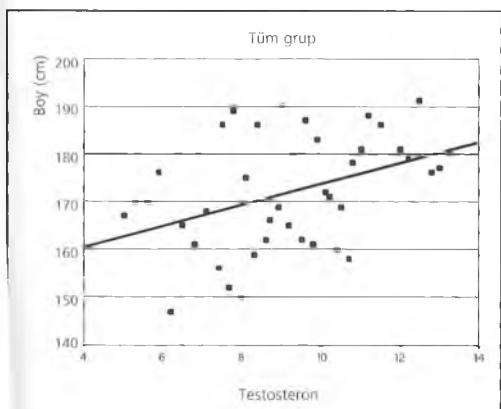
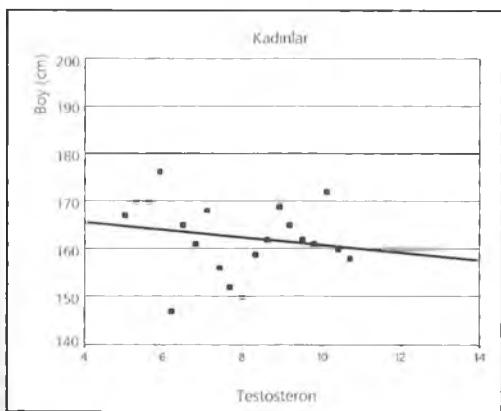
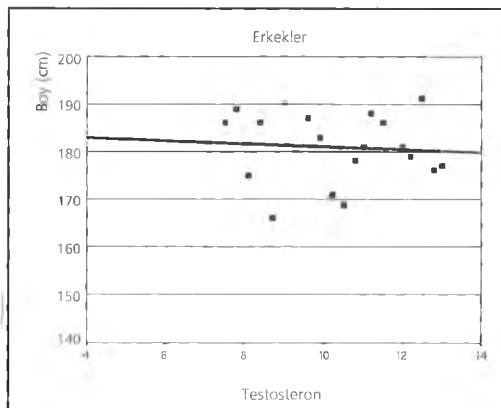


Şekil 2

Üstteki şekilde megaloblastik anemi, ortadaki şekilde demir eksikliği anemisi grubundaki Hb ve OEH verilerine ait saçılmış grafikleri ve regresyon doğruları görülmektedir. En alttaki şekilde ise iki gruptaki veriler birleştirildiğinde elde edilen regresyon doğrusu görülmektedir.

Boy ile serum testosterone düzeyi arasında ilişki olup olmadığından araştırılması amacıyla erkek ve kadın hastaları içeren bir çalışma planlanıyor. Hem erkeklerde, hem de kadınlarda boy ile testosterone düzeyi arasında ilişki olmadığı görülmüyor, ancak tüm hastalar birleştirildiğinde testosterone düzeyi arttıkça, boyun da arttığı saptanıyor.

Bu örnek de, farklı özelliklerdeki grupların birleştirilmesi ile, Örnek 71'in aksine, gerçekte ilişki olmadığı halde, ilişki olduğu sonucunun elde edilebileceğine bir örnek oluşturmaktadır. Şekil 3 incelendiğinde her iki gruptaki verilerin grafikteki dağılımının tamamen rastlantısal olduğu, yani boy ile testosterone düzeyi arasında ilişki olmadığı görülmektedir. İki grup birleştirildiğinde, iki değişken arasında kuvvetli ilişki olduğu sonucu tamamen hatalıdır. Özellikle verilerin nokta grafiğindeki dağılımı öbek öbek yerleşim gösteriyorsa, hastaların kendi içerisinde gruplaşmalar gösterdiğinden kuşkulansılmalıdır. Ancak bu örnekte olduğu gibi noktaların dağılımı hastaların altgruplardan oluşturukları yönünde bir ipucu vermese bile, ölçülen değişkenlerin özellikleri dikkate alınarak altgruplar olup olmadığı incelenmelidir.

**Şekil 3**

Örnek 72'deki çalışmada verilerin dağılım grafiği ve regresyon doğrusu. Erkeklerde boy ve testosteron düzeylerinin kadınlara göre daha yüksek olduğu, ancak her iki cinsite de boy ile testosteron düzeyi arasında ilişki olmadığı görülmektedir. İki cinsiyet grubunun birleştirilmesi ile elde edilmiş olan regresyon doğrusu ise testosteron düzeyi ile boy arasında pozitif yönde ilişki olduğu izlenimi vermektedir.

Regresyon analizinin son varsayımlı ise bağımlı değişken değerlerinin birbirlerinden bağımsız olmasıdır. Yani bir deneye ait birden fazla ölçüm bulunmamalıdır. Örneğin ultrasonografi ile ölçüm yaparak gestasyon yaşıının kesirilmeye çalışıldığı bir çalışmada, 100 gebeden değişik zamanlarda elde edilmiş 300 ölçüm varsa, bu 300 ölçümün de dahil edildiği regresyon analizi yapılamaz.

Varsayımların sağlanıp sağlanmadığı nasıl kontrol edilir?

Student t testi, varyans analizi vb. istatistik analiz yöntemlerini uygularken, elimizdeki verilerin test varsayımlarını sağlayıp sağlamadığını, analizle ilgili hesaplamalara başlamadan anlayabiliyoruz. Regresyon analizinde ise gerekli varsayımların sağlanıp sağlanmadığı çoğu kez tüm hesaplamalar bittikten sonra yapılabilir.

Eğer regresyon analizine alacağımız bağımlı ya da bağımsız değişkenlerden herhangi birinin dağılımı normal değilse, diğer varsayımların (özellikle iki değişkenin ikili normal dağılıma uyması gerektiği varsayımları) sağlanması güçleşecektir. Öte yandan her iki değişkenin dağılıminin normal olması, regresyon analizinin diğer varsayımlarının sağlanıp sağlanmadığının incelenmesini gerektirmez.

Değişkenlerin dağılımı normalse ve değişken değerleri bağımsızsa, regresyon analizine başlayabiliriz. Analize ilişkin hesaplamalar bittikten sonra da varyansların homojenliği ve ikili normal dağılım varsayımları test edilmelidir. Tüm bu incelemelerin yorumlanabilmesi, rezidüel adı verilen terimin ve ilgili kavramların iyi anlaşılabilmesine bağlıdır.

Rezidüel analizi

Tüm deneklerin verilerinin tam bir doğru üzerine oturması hemen hemen olanaksızdır. Az ya da çok regresyon doğrusundan sapmalar olacıktır. Eğer sapmalar çoksa, değişkenler arasında bağıntı bulunamayacaktır; sapmalar azsa, bağıntı bulunabilir. Bir denegin regresyon eşitliğinden hesaplanan bağımlı değişken değeri ile gerçek değeri arasındaki farka rezidüel adı verilir. Örneğin regresyon eşitliği

$$\text{Koroner arter çapı (mm)} = 5.51 - (0.0122 \times \text{Kolesterol (mg/dL)})$$

ise, kolesterol düzeyi 250 mg / dL olan bir hastanın koroner arter çapının 2.7 mm olması beklenir. Koroner arter çapı 3 mm olarak ölçülmüşse, bu hasta için rezidüel 0.3 mm'dir; yani regresyon eşitliği ile bu kadar yanılma olmuştur.

Bağımlı değişkenin kestirilmesindeki yanılma miktarı, yani rezidüeller, bağımsız değişkenin değişik değerleri için aynı olmalıdır. Örneğin kolesterol düzeyinin 100 mg/dL'den düşük olduğu ve 350 mg/dL'den yüksek olduğu durumlarda rezidüel pozitif (yani koroner arter çapı, hesaplanan değerden daha geniş) ve 100-350 mg/dL arasında negatifse (yani koroner arter çapı, hesaplanan değerden daha dar), rezidüeller kolesterol değerine bağımlıdır.

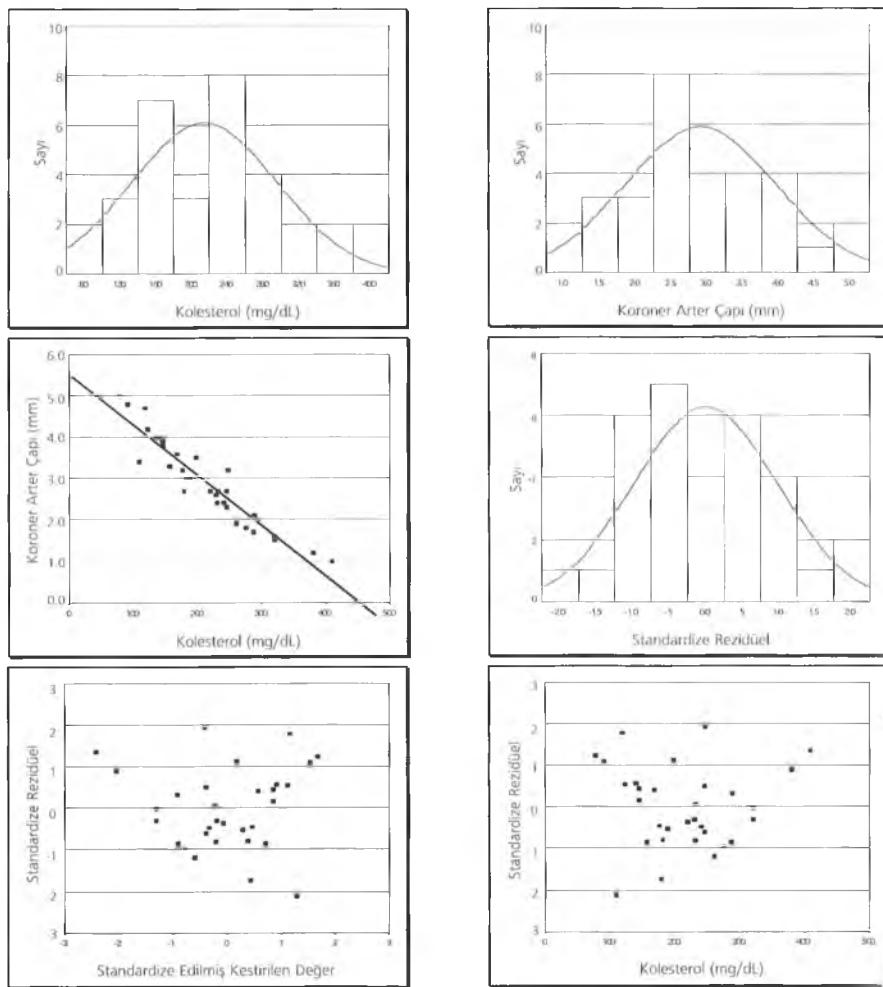
Bu durumda iki değişken arasındaki ilişkinin doğrusal olmadığından kuşkulanılmalıdır. Bunu anlamak için, yatay eksende bağımlı değişkenin kestirilen değerlerinin, dikey eksende ise rezidüellerin yer aldığı nokta grafik çizilmelidir. İlişki doğrusal ise, noktaların 0'dan çizilen yatay doğru etrafında belli bir pattern çizmeden dağılması gereklidir. Varyansların eşitliği varsayımları da benzer grafik yöntemlerle test edilebilir.

İki değişkenin dağılımı ikili normal dağılıma uyuyorsa, rezidüellerin normal dağılım göstermesi gereklidir. Buna dayanarak, bağımlı değişkenin rezidüelleri hesaplandığında, normal dağılım açısından test edilirse, regresyon analizinin en önemli varsayımlarının başında gelen ikili normal dağılım varsayımlı kontrol edilmiş olur. Regresyon analizi varsayımlarının hangi yöntemlerle incelendiği Tablo 1'de listelenmiştir.

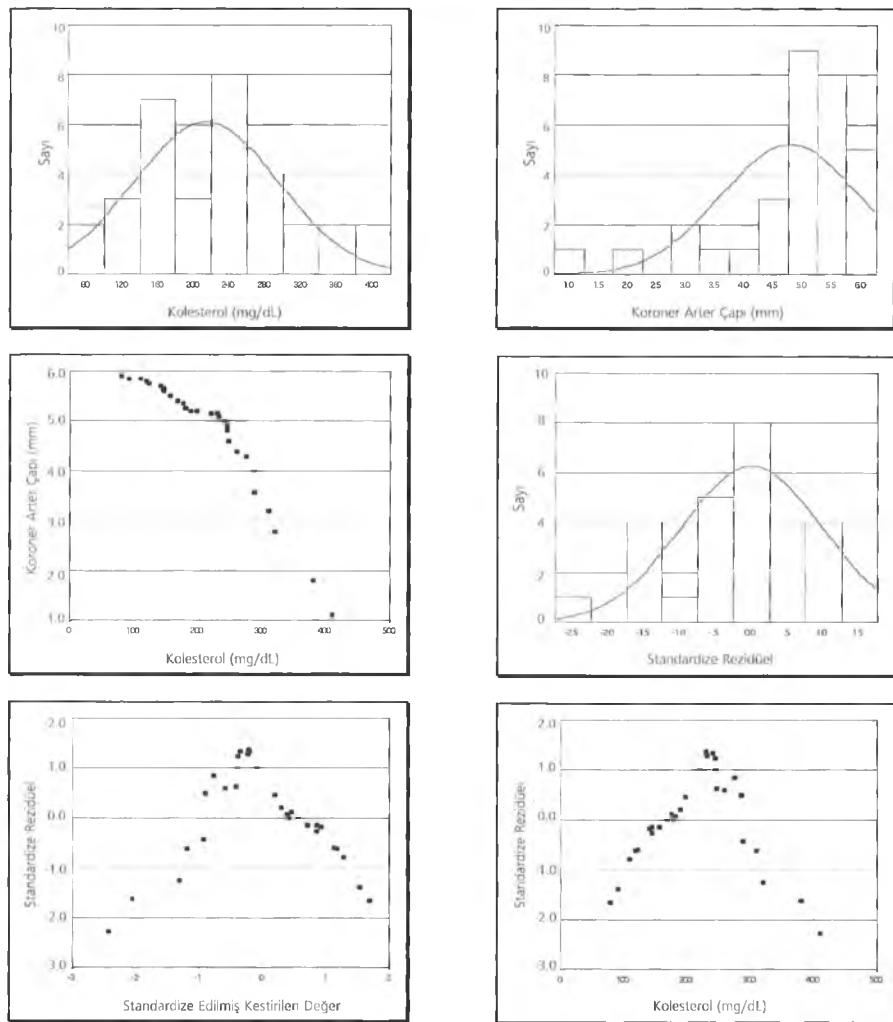
Tablo 1 Regresyon analizi varsayımlarının sağlanıp sağlanmadığına ilişkin yöntemler.

Yöntem	Hangi varsayımlar için?	Açıklama
Bağımlı ve bağımsız değişkenlerin normal dağıldığını gösterin.	İkili normal dağılım	Her iki değişkenin dağılımının normal olması gereklidir. Ama tek başına ikili normal dağılım varsa yarının sağlandığı anlamına gelmez.
Rezidüllerin normal dağıldığını gösterin.	İkili normal dağılım	İkili normal dağılım varsayımlarının sağlandığını göstermek için rezidüllerin normal dağıldığının gösterilmesi yeterlidir.
"Kestirilen değer – rezidüel" ya da "bağımsız değişken değeri – rezidüel" nokta grafiklerinde belirli bir pattern olmadığını görün.	Doğrusallık	Bağımlı ya da bağımsız değişkenin her değeri için rezidüllerin dağılımı benzer olmalıdır.
"Kestirilen değer – rezidüel" ya da "bağımsız değişken değeri – rezidüel" nokta grafiklerinde rezidüllerin dağılımının kestirilen değer ya da bağımsız değişkenin her değeri için benzer olduğunu görün.	Varyansların homojenliği	Bağımlı ya da bağımsız değişkenin her değeri için rezidüllerin dağılımı benzer olmalıdır.

Bölüm 13'te verilen Örnek 67'deki çalışmaya ilişkin rezidüel analizi yapıldığında Şekil 4'deki grafikler elde edilmiştir. Başka bir çalışmaya ilişkin rezidüel analizi grafikleri ise Şekil 5'de görülmektedir. İki şekildeki grafikleri karşılaştırınız.



Şekil 4. Örnek 67'deki çalışmada kolesterol - koroner arter çapı regresyon analizi için varsayımların sağlanıp sağlanmadığını incelemek için kullanılan grafikler. Üst sırada iki şekilde kolesterol ve koroner arter çapının histogramları görülmektedir. Orta sırada soldaki şekilde regresyon doğrusu ve kolesterol - koroner arter çapı saçılım grafiği görülmektedir. Diğer şekillerde rezidüllerin histogramı, standartize kestirilen değer (standartize koroner arter çapı) - rezidüel saçılım grafiği ve bağımsız değişken (kolesterol) - rezidüel saçılım grafiği görülmektedir. Rezidüllerin histogramları dan rezidüllerin dağılımının normale yakın olması ve en alttaki iki grafikteki noktaların belirli bir pattern çizmeden dağılması, regresyon analizi varsayımlarının sağlandığını göstermektedir.



Şekil 5 Örnek 67'deki çalışmaya benzer başka bir çalışmada kolesterol - koroner arter çapı regresyon analizi için varsayımların sağlanıp sağlanmadığını incelemek için kullanılan grafikler. Üst sıradaki iki şekilde kolesterol ve koroner arter çapının histogramları görülmektedir. Orta sıradaki şekilde regresyon doğrusu ve kolesterol - koroner arter çapı saçılım grafiği görülmektedir. Bu grafikte iki değişken arasındaki ilişkinin doğrusal olmadığı izlenimi almaktadır. Diğer şekillerde rezidüllerin histogramı, standartize kestirilen değer (standartize koroner arter çapı) - rezidüel saçılım grafiği ve bağımsız değişken (kolesterol) - rezidüel saçılım grafiği görülmektedir. Rezidüllerin histogramundan rezidüllerin normal dağılım göstermediği görülmektedir. En alttaki iki grafikteki noktaların rastlantısal dağılmaması, yani belirli bir pattern çizerek dağılması da, regresyon analizi varsayımlarının sağlanmadığını göstermektedir.

Bağımsız değişken sayısı birden fazla ise

Koroner arter çapının daralmasına neden olan aterosklerozun tek risk etkeni hiperlipidemi değildir. Hipertansiyon, diabetes mellitus, sigara içimi, erkek cinsiyet aterosklerozu hızlandıran diğer etkenler arasında yer alır. O halde daha önceki paragraflarda kolesterol düzeyinden koroner arter çapının kestirildiği eşitlikte bu değişkenlerle ilgili veriler kullanılmadığına göre, hatalar ya da eksiklikler olması olasıdır.

Burada bağımlı değişken koroner arter çapı ise, bağımsız ya da belirleyici değişkenler, yukarıda saydığımız risk etkenleridir. Bu etkenlerden cinsiyet dışındaki (kan basıncı, kan glukoz düzeyi, gündeme içilen sigara sayısı vb.) sayısal değişkenler olduğu halde, cinsiyet ikili nominal değişkendir.

Bağıntı ve regresyon analizinde hem bağımlı, hem de bağımsız değişkenler, sayısal değişkenler olmak zorunda olduğuna göre, yukarıdaki paragraflardaki regresyon eşitliklerinde kolesterol yerine, cinsiyet dışındaki değişkenlerden her birini koyarak regresyon eşitlikleri bulalım. (Cinsiyet gibi ikili değişkenler de bazen bağımsız değişken olarak işleme alınmaktadır. Ama klasik olarak regresyon analizinde bağımlı ve bağımsız değişkenler sayısal olmalıdır).

Bağımsız değişkenlerin katsayıları ve önemlilik derecesine karşılık gelen p değerleri şöyle olsun:

$$\text{Koroner arter çapı (mm)} = 5.51 - (0.012 \times \text{Kolesterol (mg/dL)}) \quad p = 0.001$$

$$\text{Koroner arter çapı (mm)} = 4.85 - (0.014 \times \text{Sistolik KB (mmHg)}) \quad p = 0.048$$

$$\text{Koroner arter çapı (mm)} = 3.34 - (0.003 \times \text{Glukoz (mg/dL)}) \quad p = 0.38$$

$$\text{Koroner arter çapı (mm)} = 3.40 - (0.046 \times \text{Sigara (adet/gün)}) \quad p = 0.06$$

Göründüğü gibi dört bağımsız değişkenin de regresyon katsayılarının işaretleri negatif olduğu için hepsinin de artan değerleri ile koroner arter çapının daraldığı anlaşılmaktadır. Kan glukoz düzeyinin ve gündeme içilen sigara miktarının regresyon katsayısına ilişkin p değerleri >0.05 olduğu için, bu değişkenlerin, koroner arter çapını belirlemek için uygun olmadığı, başka bir deyişle kan glukoz düzeyi ve gündeme içilen sigara miktarı ile koroner arter çapı arasında bağıntı olmadığı söylenebilir.

Hiperlipidemi, hipertansiyon, hiperglisemi ve sigaranın aterosklerozu hızlandırıcı etkileri hem doğrudan, hem de diğer faktörlerin etkilerini artırarak olabilir. Oysa yukarıdaki eşitliklerde her bağımsız değişkenin diğer bağımsız değişkenlerden etkilenmediği varsayılmıştır. Bağımsız değişkenlerin hem doğrudan etkileri, hem de diğer bağımsız değişkenler üzerinden olan etkiler nedeniyle, tüm bağımsız değişkenleri aynı eşitlikte koymakta yarar vardır.

Her bağımsız değişkeni içeren ayrı regresyon eşitliklerinin kullanılması, regresyon analizinin amacı olan kestirim yapabilme açısından da uygun değildir. Bunu şu örnekle açıklayalım: Kan kolesterol düzeyi 120 mg/dL, sistolik kan basıncı 140 mmHg, kan glukoz düzeyi 110 mg/dL olan ve günde 10 adet sigara içen birinin koroner arter çapı kaç mm'dır?

Bu soruyu yanıtlamak için yukarıdaki eşitlikleri kullanalım. Koroner arter çapı kolesterol düzeyine ilişkin eşitlikten 4.1 mm, sistolik kan basıncına ilişkin eşitlikten 2.9 mm, glukoz düzeyine ilişkin eşitlikten 3.0 mm ve sigara ya ilişkili eşitlikten 3.0 mm olarak hesaplanabilir. Bunlardan hangisi doğrudur? Amaç, iyi kestirim yapmak olduğu halde, 2.9 ile 4.1 mm arasında değişen değerler elde ettik. Bunun çözümü, her bağımsız değişkeni içeren regresyon eşitlikleri kullanmak yerine, tüm bağımsız değişkenleri içeren tek bir regresyon eşitliği kullanmaktır.

Çoklu regresyon analizi adı verilen teknikteki regresyon eşitliği, " $Y = \beta_0 + \beta_1 X$ " şeklindeki regresyon eşitliğinin genelleştirilmiş şeklidir. Değişken sayısı "n" olduğunda, çoklu regresyon eşitliği aşağıdaki şekildedir.

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \dots + \beta_n \cdot X_n$$

Çoklu regresyon eşitliğindeki β katsayılarına kısmi regresyon katsayıları adı verilir.

Şimdi dört kez tek bağımsız değişkenli regresyon analizi yapmak yerine, yukarıdaki çoklu regresyon yöntemini kullanırsak, aşağıdaki regresyon eşitliğini ve her bağımsız değişkenin regresyon katsayısına karşılık gelen p değerlerini elde ederiz.

Koroner arter çapı (mm) = 5.71	- 0.012 X Kolesterol (mg/dL)	p = 0.001
	- 0.003 X Sistolik KB (mmHg)	p = 0.31
	+ 0.0012 X Glukoz (mg/dL)	p = 0.38
	- 0.0004 X Sigara (adet/gün)	p = 0.97

Bu eşitliği inceleyeceğiz olursak, ilginç durumlara rastlıyoruz. Öncelikle kısmi regresyon katsayılarının daha önceki değerlerine göre değiştigini görüyoruz. Örneğin sigara sayısının regresyon katsayısının mutlak değeri küçüldü. Tek bağımsız değişken olarak alındığında regresyon katsayısı -0.046 iken, bu eşitlikte -0.0004 olarak değişti. Öte yandan glukozun regresyon katsayısının işaretini değiştirdi. Yani glukoz değeri arttıkça, koroner arter çapı da artıyor. Bu hiç de beklememişti bir bulgu. Çünkü kan glukoz düzeyi artınca koroner arter çapı artıyor gibi görünüyor. Fakat regresyon katsayısına karşılık ge-

len p değerine bakarsanız, 0,05'ten hayli büyük ($p=0,38$). Bu p değerinin, regresyon katsayısının 0'a eşit olduğu hipotezinin test edilmesinden elde edildiğini hatırlarsanız, iş açıklığa kavuşuyor. Yani işaretti pozitif ya da negatif olsun, aslında glukozun regresyon katsayısı 0'dan farklı değil. Bu nedenle kan glukoz düzeyi artınca, koroner arter çapının arttığını söylemek olanaksızdır.

Yine de herşey açık gözükmüyor. Çünkü, daha önceki paragraflarda tek değişkenli regresyon analizi yapıldığında, sistolik kan basıncının katsayısı negatifti ve p değeri de 0,05'ten küçüktü. Hipertansiyonla koroner arter çapı arasında negatif bir bağıntı bulmuştuk. Çoklu regresyon analizi yapınca, böyle bir bağıntı yok gibi görülmüyor. Bu ne anlama gelmektedir?

Bunun anlamı, bağımsız değişkenlerin birbirlerini etkilemeleri, yani birbirlerinden bağımsız olmamaları ile açıklanabilir. Sistolik kan basıncı daha yüksek olanların aynı zamanda kan kolesterol düzeyleri de yüksekse, hipertansiyonun koroner arter çapında daralmaya neden olması, aslında doğrudan kendi etkilerine bağlı değil, kolesterol düzeyinin daha yüksek olmasına bağlı olabilir. İşte çoklu regresyon analizi yerine tek değişkenli regresyon analizi uygulanması ile bu ilişki gözden kaçmıştır. Şekil 6'da tüm değişkenler arasındaki bağıntı katsayıları ve p değerleri verilmiştir. Kolesterol ile sigara ve sistolik kan basıncı arasındaki bağıntı katsayılarının oldukça büyük olduğu görülmektedir.

Correlations				
		KOLESTEROL (mg/dL)	SKB (mmHg)	GLUKOZ (mg/dL)
SKB (mmHg)	Pearson Correlation	.323	1.000	.261
	Sig. (2-tailed)	.077		.156
GLUKOZ (mg/dL)	Pearson Correlation	.219	.261	1.000
	Sig. (2-tailed)	.237	.156	
SIGARA (adet/gün)	Pearson Correlation	.351	.136	-.080
	Sig. (2-tailed)	.053	.464	.668

Şekil 6 Örnek 67'deki çalışmada değişkenler arasındaki bağıntı katsayıları.

Çoklu regresyon analizi ile bağımlı değişken üzerinde bağımsız değişkenlerin etkileri arıtılmaktadır. Çoklu regresyon analizinin tek yararı bu degildi

Daha önce dört regresyon eşitliği ile koroner arter çapını kestirmeye çalıştığımızda, çok farklı dört sonuç elde etmiştık. İşte çoklu regresyon eşitliğini kullanarak, en doğru kestirimini yapmak olası olacaktır. Kan kolesterol düzeyi 120 mg / dL, sistolik kan basıncı 140 mmHg, kan glukoz düzeyi 110 mg / dL olan ve günde 10 adet sigara içen birinin koroner arter çapını yukarıdaki eşitlikten şöyle hesaplayabiliriz:

$$\begin{aligned}\text{Koroner arter çapı (mm)} &= \\ &= 5.71 - (0.012 \times 120) - (0.003 \times 140) + (0.0012 \times 110) - (0.0004 \times 10) \\ &= 5.71 - 1.44 - 0.42 + 0.13 - 0.004 = 4.0 \text{ mm}\end{aligned}$$

REGRESYON EŞİTLİĞİNE ALINACAK BAĞIMSIZ DEĞİŞKENLERİ NASIL SEÇELİM?

Regresyon analizinde birden çok değişkenin bağımsız değişken olarak alınabileğini gördük. Peki regresyon eşitliğine alınabilecek bağımsız değişken sayısı ne kadar olmalıdır? Bu sorunun kesin bir yanıtı yok. Ama denek sayısının, bağımsız değişken sayısının 10 katı kadar olması genellikle kabul görmektedir. Yukarıdaki örnekte dört bağımsız değişkeni de regresyon eşitliğine alabilmek için denek sayısının 40 civarında olması istenir.

Bir diğer sorun ise, kısmi regresyon katsayılarının önemlilik p değeri 0.05'ten büyük olan değişkenleri eşitlikte tutmanın yararlı olup olmadığına ilişkindir. Gerçekten de madem ki sigara, sistolik kan basıncı ve kan glukoz düzeyi, koroner arter çapını kestirmek için yeterince önemli değil, o halde bu değişkenleri neden regresyon eşitliğinin içinde tutalım?

Regresyon eşitliğinde hangi değişkenlerin bulunacağı, hangilerinin bulunmayacağıın belirlenmesinde çeşitli yöntemler vardır. Bu yöntemlerin hiçbirinin diğerine belirgin üstünlüğü yoktur. Sık kullanılan üç değişken seçme ve eleme yöntemi ileri doğru seçme (*forward selection*), eleme (*backward elimination*) ve adım adım seçme ve eleme (*stepwise selection*) yöntemleridir.

Forward selection

Bu yöntemde başlangıçta regresyon eşitliğinde hiç bağımsız değişken yoktur. İlk adımda bağımsız değişkenlerin bağımlı değişkenle aralarındaki bağıntı katsayılarına bakılarak, mutlak değeri en yüksek olan regresyon eşitliğine alınır. Eğer regresyon eşitliğindeki regresyon katsayısına karşılık gelen p değeri önemli ise, bu değişken eşitlikte kalır. İkinci adımda henüz eşitliğe alınmamış değişkenlerden yine bağıntı katsayısının mutlak değeri en büyük

olan eşitlige alınır ve böylece devam eder. Eşitlige alınmamış değişkenlerden bağıntı katsayısı en büyük olan bile eşitlige alınamazsa, işlem durur ve o ana kadar eşitlige alınmış değişkenlerle regresyon eşitliği kurulmuş olur. Yukarıdaki örnekte değişken seçiminde *forward selection* yöntemi uygulandığında, kolesterol değişkeni eşitlige alındıktan sonra, işlem durmuştur. Çünkü sigara, sistolik kan basıncı ve glukoz değişkenlerinin eşitlige alındıkları takdirde regresyon katsayılarının p değerleri sırasıyla 0.82, 0.37, 0.48 olacaktır, yani 0.05'ten büyüktür.

Backward elimination

Bu yöntemde *forward selection* yönteminin tersine, başlangıçta tüm bağımsız değişkenler regresyon eşitliğine dahil edilmiştir. Bu kez, p değeri 0.05'ten büyük olan değişkenler sırasıyla teker teker elenmeye çalışılır. Regresyon eşitliğinde kalan tüm değişkenlerin p değerleri 0.05'ten küçükse, işlem durur. Bazen, *backward elimination* yönteminde değişkenin elenmesi için kriter olarak p değerinin 0.05 yerine, 0.10'dan büyük olması arandığı için, *forward selection* yöntemine göre daha fazla sayıda değişken, regresyon eşitliğinde kalabilir. Koroner arter çapı örneğinde bu yöntemle de kolesterol dışındaki değişkenler eşitlikten elenmiştir.

Stepwise selection

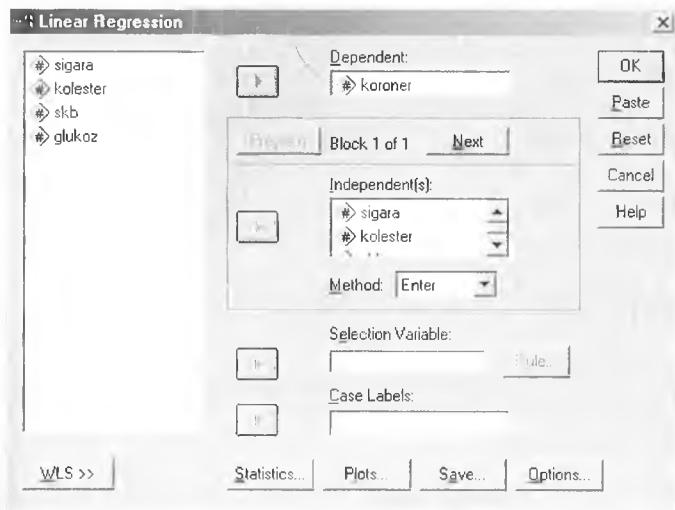
Bu yöntem, *forward selection* ve *backward elimination* yöntemlerinin karışımı şeklindedir. Değişken seçimi önce *forward* olarak başlar. Ancak regresyon eşitliğine her yeni değişken alındığında, bu kez eşitlikteki tüm değişkenlerin *backward* olarak elenme kriterlerine uygun olup olmadığı araştırılır. Eğer eşitlikte kalan değişkenler elenemiyorsa ve eşitlik dışında kalan değişkenler seçilemiyorsa işlem durur.

Bu üç değişken seçme yöntemi ile genellikle aynı sonuç elde edilir.

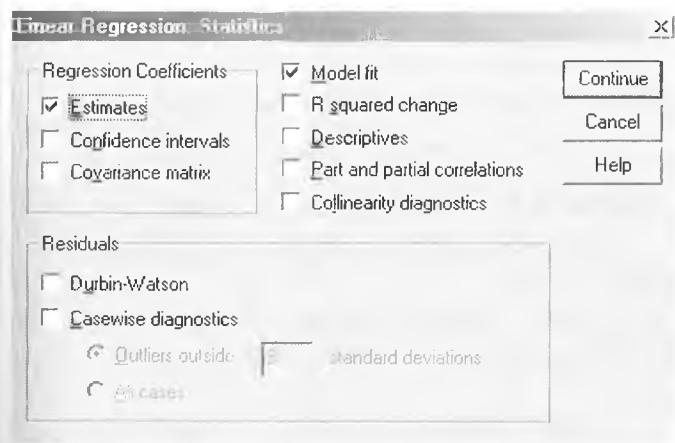
SPSS'TE REGRESYON ANALİZİ NASIL YAPILIR?

Analyze >> Regression >> Linear

Yukarıdaki seçimle **Linear Regression** diyalog kutusu açılır (Şekil 7). Sol daki listede sayısal olarak tanımlanmış değişkenler vardır. *Dependent* kutusu na bağımlı değişken, *Independent(s)* kutusuna bağımsız değişken(ler) seçilip OK'lerseniz, tüm bağımsız değişkenlerin dahil edildiği modelle regresyon analizi yapılır. *Method* kutusundan istediğiniz bağımsız değişken seçimi yöntemi seçebilirsiniz.



Şekil 7
SPSS'te "Linear Regression" diyalog kutusu.



Şekil 8
SPSS'te "Linear Regression" diyalog kutusundan "Statistics" altmenüsü.

Statistics altmenüsünde *Estimates* ve *Model fit* önceden seçilmiştir (Şekil 8). Bu seçeneklerle regresyon katsayıları, bunlara karşılık gelen p değerleri, modelin "r" ve "çoklu r" değerleri gibi bazı değerler hesaplanabilir. Bu altmenu'deki diğer seçeneklerle regresyon katsayılarının güven sınırları ve bazı diğer istatistikler de yapılabilir. Önceden belirlenmiş seçeneklerle regresyon analizi yapıldığında Şekil 9'daki çıktı elde edilecektir.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.945 ^a	.893	.876	.369

a. Predictors: (Constant), GLUKOZ, SIGARA, SKB, KOLESTER

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	29.448	4	7.362	54.190	.000 ^a
	Residual	3.532	26	.136		
	Total	32.980	30			

a. Predictors: (Constant), GLUKOZ, SIGARA, SKB, KOLESTER

b. Dependent Variable: KORONER

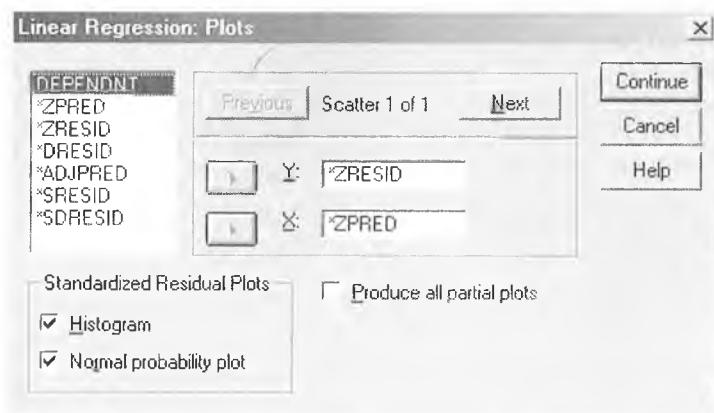
Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.
	B	Std. Error			
1	(Constant)	5.713	.372	15.353	.000
	SIGARA	-3.60E-04	.009	-.038	.970
	KOLESTER	-1.20E-02	.001	-.930	.000
	SKB	-2.91E-03	.003	-.073	.301
	GLUKOZ	1.197E-03	.001	.061	.383

a. Dependent Variable: KORONER

Şekil 9 SPSS'te regresyon analizi sonuçları.

Plots altmenüsünden varsayımların sağlanıp sağlanmadığını incelemeyi yarayacak grafikler ve rezidüel analizi yapılabılır (Şekil 10). Bu menüde sol daki listedeki değişkenleri "Y" ya da "X" eksene yerleştirerek, istediğim nokta grafikleri çizdirebilirsiniz.

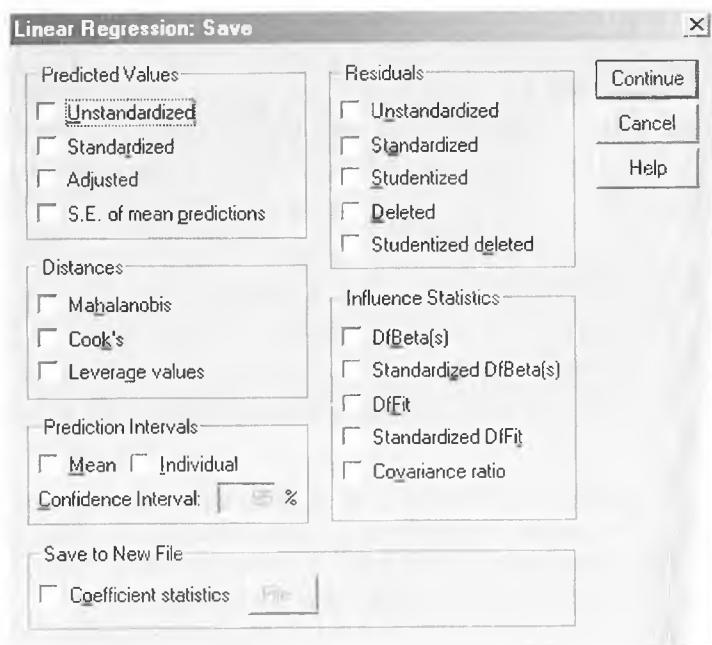


Şekil 10
SPSS'te "Linear Regression" diyalog kutusundan "Plots" altmenüsü.

"*ZPRED" ve "*ZRESID" ile gösterilenler, sırasıyla standardize edilmiş kestirilen değer ve standardize edilmiş rezidüeldir. "SRESID" ile gösterilen Studentize rezidüel de standarize edilmiş rezidüel yerine kullanılabilir. İkisi arasında genellikle önemli bir fark yoktur. Studentize rezidüelin kullanılmasının daha doğru olduğu öne sürülmektedir. Şekil 4 ve 5'de kestirilen değer - rezidüel grafikleri, bu menüden Y eksenine "*ZRESID" ve X eksenine "*ZPRED" seçilerek çizdirilmiştir.

Yine *Plots* altmenüsünden standardize rezidüellerin histogramı ve normal olasılık grafiği çizdirilebilir. Şekil 4 ve 5'deki histogramlar bu menü yardımı ile çizdirilmiştir.

Save altmenüsü yardımı ile her deneğe ilişkin kestirilen değerler, kestirim sınırları ve rezidüeller o anda açık olan veri dosyasına kaydedilebilir (Şekil 11). Şekil 4 ve 5'deki bağımsız değişken - rezidüel grafikleri, bu menüden kaydedilen rezidüelleri kullanarak "Graphs > Scatter" menüsünden çizdirilmiştir.



Şekil 11
SPSS'te "Linear Regression" diyalog kutusundan "Save" alt-menüsü.

bölüm 15

SAĞKALIM ANALİZİ

Sansürlü değişken nedir?

Yaşam tablosu analizi

Yaşam tablosu analizinin varsayımları

Kaplan-Meier sağkalım analizi

Sağkalım analizinde risk fonksiyonu

İki sağkalım eğrisinin karşılaştırılması

Gehan (genelleştirilmiş Wilcoxon) testi

Logrank testi

Cox regresyon analizi

SPSS'te sağkalım analizi nasıl yapılır?

Yaşam tablosu

Kaplan-Meier sağkalım analizi

Cox regresyon analizi

Tıpta yeni bir ilaç ya da tedavi yönteminin bireyler üzerine kısa süreli etkileri dışında mortaliteye etkilerinin incelendiği araştırmalar da kullanılmaktadır. Ayrıca farklı özellikleri olan hasta gruplarındaki yaşam süreleri karşılaştırılarak, prognoza etki eden etkenlerin incelenmesi de biyolojik bilimlerin ilgi alanlarının başlarında gelmektedir. Farklı tedavi yöntemleri uygulanan hastalarda yaşam sürelerinin karşılaştırılması, onkologlar ve hematologlar başta olmak üzere kanserle uğraşan araştırmacıların önemli uğraş konusudur.

SANSÜRLÜ DEĞİŞKEN NEDİR?

Bu tür çalışmalarında, çalışma grubuna alınan tüm hastalar, öldükleri güne kadar izlendikten sonra farklı gruppardaki hastaların yaşam sürelerinin ortalaması ya da ortancalarının bilinen testlerle karşılaştırılması ile amaca ulaşılabilir. Ama bu süre çok uzun olabileceği gibi başarıyla tedavi edilmiş hastalar, kanser dışı nedenlerle de ölebilirler.

Örneğin farklı tedavi yöntemi uygulanan her birinde 30'ar hasta bulunan

üç grup olduğunu düşünelim. Grplarda 5 ile 15 hastanın en az 10 yıl yaşayacağını, hatta 23 hastanın da belki 20-30 yıl daha yaşayacağını varsayıyalım. Çalışmaya başladıkten 10 yıl sonra veriler değerlendirildiğinde o tarihe kadar ölmüş olan hastaların yaşam süresi (yaşam süresi = izlem süresi) net olarak bilinmekte beraber, halen yaşamakta olan hastaların bir gün sonra, ya da 50 yıl sonra ölüp ölmeyecekleri belli değildir (yaşam süresi > izlem süresi).

O halde bu hastaların yaşam süreleri net olarak bilinmemektedir. Öte yandan bazı hastalar tedaviden sonra bir süre sonra kontrole gelip, daha sonra izlemeden çıkış olabilirler. Ölüm ölümedikleri, eğer öldülerse, ne zaman öldükleri belli değildir. Bu hastaların da yaşam süreleri net olarak bilinmemektedir (yaşam süresi > izlem süresi). Yaşayan ya da izlemeden çıkış hastalara ait yaşam sürelerinin bir şekilde incelemeye katılması gereklidir. İşte bu örnekteki izlem süresi gibi son durumun belirli olmadığı değişkenlere sansürülü değişkenler (*censored variables*) adı verilir. Bu nedenle yaşam süresiyle ilgili çalışmalarında özel teknikler kullanılması gereklidir. Bunlardan biri Örnek 73'de görülebilir.

ÖRNEK 73

Aplastik anemide androjen ve immünosupresif tedavinin yaşam süresine etkilerinin incelenmesi amacıyla yapılacak bir çalışmaya 1 Ocak 1999 tarihinde başlanıyor ve yeni tanı konulan hastalar, uygun randomizasyon yöntemleri ile ya androjen, ya da immünosupresif tedavi verilecek gruplara alınıyor. Çalışmaya son olarak 31 Aralık 2000 tarihinde hasta alınacağı ve 1 Şubat 2001 tarihine kadar hastalar izlendikten sonra bu tarihte toplanan verilerle analize başlanacağı planlanmıştır.

1 Ocak 1999 tarihinde görülen ilk hasta androjen grubuna, 9 Ocak 1999 tarihinde görülen ikinci hasta immuno-supresif tedavi grubuna, ..., 24 Aralık 2000 tarihinde görülen 59. hasta androjen grubuna ve 30 Aralık 2000 tarihinde görülen 60. hasta immuno-supresif tedavi grubuna alınmış ve 60. hasta ile birlikte, planlandığı gibi çalışmaya yeni hasta alınması durdurulmuş olsun. 1 Şubat 2001 tarihinde elimizde her grupta 30'ardan toplam 60 hastaya ait izlem süreleri ve bu sürenin sonunda ne durumda oldukları, yani sağ mı, yoksa ölmüş mü olduklarına ait bilgi vardır. İzlem sürelerinin sonunda, yani en son kontrole geldiklerinde ya da kendilerinden haber alındığında hastalar ya ölmüş olacaklardır, ya da sağ olacaklardır. Bu arada hastaların bir kısmı beklenen kontrollerine gelmemiş olabilirler. Örneğin birinci hastanın en son 1

Şubat 2000 tarihindeki kontrolde sağ olduğu bilindikten sonra kontrole gelmediğini varsayıyalım. Bu hastanın izlem süresi 13 aydır, ama bu süre sonunda hasta hala sağ olduğu için yaşam süresi bilinmemektedir. Hasta belki son kontrolden bir ay sonra öldüğü için kontrole gelmemiştir (bu durumda yaşam süresi 14 ay olacaktır), ya da sağ olduğu halde kontrole gelmemiştir ve belki 10 yıl daha yaşayacaktır (bu durumda yaşam süresi çok daha uzun olacaktır). Öte yandan çalışmaya son alınan hastalar ise kontrollere düzenli gelseler ve son kontrollerinde sağ oldukları biliniyor olsa bile, analize geçilecek tarihte, yani Şubat 2001'de en fazla 1-2 aydır izleniyor olacaklardır. Bu hastaların da yaşam süresi henüz belirli değildir.

O halde ölmüş olan hastalar için "izlem süresi = yaşam süresi" olduğu halde son görüldüklerinde sağ olan hastalar için "izlem süresi < yaşam süresi"dir. Bu örnekte sağ kalan hastaların yaşam süresi, henüz netleşmemiş olduğu için sansörlü bir değişkendir. Sansörlü değişkenlerin karşılaştırılmasında standart istatistik analiz yöntemlerinin uygulanması hatalıdır. Ortalamaların karşılaştırılması gibi standart yöntemlerin nasıl hatalı olacağını Örnek 74'de inceleyelim.

ÖRNEK 74

Kanserli 30 hasta 10'arlı üç gruba ayrılıyor ve bir gruba yalnız kemoterapi, ikinci gruba yalnız radyoterapi ve üçüncü gruba ise kemoterapi + radyoterapi birlikte veriliyor. Çalışma bittiğinde ya da izlemden çıkmadan önceki en son kontrolünde sağ olanlar, izlem sürelerinden daha fazla yaşam süreleri olması olasılığı açısından aynı gruba alınabilirler. Bu nedenle her iki durumdaki hastaları birbirinden ayırmaya gerek yoktur. Hastaların durumları Tablo 1 ve 2'de listelenmiştir.

Tablo 1'de görüldüğü gibi üçüncü grupta ortalama ve ortanca izlem süresi (yaşam süresi değil !) diğer iki gruba göre daha kısadır (Grplarda henüz yaşamakta olan hastalar olduğu ve bu hastalar yaşamaya devam edecekleri için ortalama izlem süreleri, ortalama yaşam sürelerinden daha kısadır).

Kemoterapi ve radyoterapinin birlikte verilmesi ile sağlam süresi diğer grplara göre daha kısalmış gibi görülmektedir. Oysa üçüncü grupta 10 hastadan biri ölmüşken, diğer iki grupta dörder hastanın öldüğü görülmektedir. Bu gelişkinin nedenini anlamak için üçüncü grupta yaşayan 9 hastayı ele alalım. İzlem süresi 12 ayı bulan 5 hasta hala sağdır ve belki daha çok uzun süre yaşamaya devam edeceklerdir. Oysa diğer iki grupta izlem süresi 12 ayı

bulan hastaların yaklaşık yarısı ölmüştür. Bu nedenle bu grplardaki diğer hastaların da 12 ayı geçtikten sonra ölme olasılığı yüksek gibi görülmektedir.

Tablo 1 Örnek 74'teki çalışma henüz sonlanmadan yapılan bir ara değerlendirme sonuçları.

No	1. Grup KT ¹		2. Grup RT ²		3. Grup KT + RT	
	İzlem süresi	Son durum	İzlem süresi	Son durum	İzlem süresi	Son durum
1	3	Yaşıyor	5	Yaşıyor	5	Yaşıyor
2	5	Yaşıyor	6	Yaşıyor	8	Öldü
3	7	Öldü	7	Yaşıyor	9	Yaşıyor
4	10	Yaşıyor	12	Öldü	9	Yaşıyor
5	12	Yaşıyor	12	Yaşıyor	10	Yaşıyor
6	12	Yaşıyor	12	Yaşıyor	12	Yaşıyor
7	14	Öldü	14	Öldü	12	Yaşıyor
8	24	Yaşıyor	15	Öldü	14	Yaşıyor
9	24	Öldü	16	Öldü	16	Yaşıyor
10	26	Öldü	18	Yaşıyor	20	Yaşıyor
Ortalama:		14	Ortalama:	12	Ortalama:	11.5
Ortanca:		12	Ortanca:	12	Ortanca:	11

1 Kemoterapi

2 Radyoterapi

Şimdi bu çalışmanın 7 yıl süregünü ve üç gruptaki hastaların da öldüğüünü varsayıyalım. Hastalar öldüğü için izlem süreleri, yaşam sürelerine eşit olacaktır. Buna göre rakamları değiştirirsek, aşağıdaki Tablo 2'yi elde ederiz. Bu tablodan üçüncü grupta izlem süresinin (= yaşam süresi) belirgin olarak uzun olduğu görülmektedir. Grplardaki hastaların hepsi öldüğü için artık izlem süresi, yaşam süresine eşittir. Yani izlem süresi artık sansürlü değişken değildir. Bu örnektenden anlaşıldığı gibi ortalaması yaşam süreleri, verilerin ne zaman analiz edildiğine bağlıdır. Ne kadar erken analiz yapılrsa, ortalaması yaşam süreleri gerçekten o kadar uzak ve yaniltıcıdır. Tüm hastalar ölene kadar beklenmekten sonra analiz yapılrsa, ortalaması yaşam süresi gerçeği yansıtır.

Tablo 2 Örnek 74'teki çalışma sonliğinde yapılan değerlendirmenin sonuçları.

No	1. Grup KT ¹		2. Grup RT ²		3. Grup KT + RT	
	İzlem süresi	Son durum	İzlem süresi	Son durum	İzlem süresi	Son durum
1	9	Öldü	8	Öldü	24	Öldü
2	12	Öldü	12	Öldü	8	Öldü
3	7	Öldü	8	Öldü	34	Öldü
4	12	Öldü	12	Öldü	12	Öldü
5	15	Öldü	15	Öldü	36	Öldü
6	15	Öldü	28	Öldü	48	Öldü
7	14	Öldü	14	Öldü	50	Öldü
8	28	Öldü	15	Öldü	18	Öldü
9	24	Öldü	16	Öldü	20	Öldü
10	26	Öldü	36	Öldü	66	Öldü
Ortalama:		16	Ortalama:		16	Ortalama:
Ortanca:		14.5	Ortanca:		14.5	Ortanca:
						29

¹ Kemoterapi² Radyoterapi

Son örnekteki gibi tüm hastaların ölmesini beklemek için zamanın geçmesini beklemek, özellikle kanser tedavisi gibi hızlı gelişen alanlarda hiç de pratik değildir. Bu nedenle yaşam süresi analizinin bilinmesi çok yararlıdır.

Başlıca iki tip sağkalım analiz yöntemi vardır: Yaşam tablosu analizi ve Kaplan-Meier sağkalım analizi.

YAŞAM TABLOSU ANALİZİ

Sağkalım süresi analizine yaşam tablosu (*life-table*) adı verilen tablonun oluşturulması ile başlanır. Yaşam tablosu yapmak için hastaların izlem süreleri ve bu sürenin sonundaki durum (*status*) bilinmelidir. İzlem süresinin başlangıcı iyi tanımlanmalıdır. İlk hastalık bulgularının başlaması, hastalık tanısının konulması, tedaviye başlanması, ameliyat yapılması vb. olayların tarihleri izlem sürelerinin başlangıcı olarak tanımlanabilir. Son durum seçenekleri "öldü", "yaşıyor" olarak tanımlanabileceği gibi, "remisyonda", "relaps" olarak da tanımlanabilir. Yani "son durum" bir "olay"ın (*event*) gerçekleşmesi ya

da gerçekleşmemesi olarak tanımlanır. Yukarıdaki cümlede tanımlanan "son durum"u belirleyen "olay" birinci örnekte "ölüm", ikinci örnekte "relaps"tır. Sonuç olarak izlem süresi, "olay" a kadar geçen süredir (*time-to-event data*).

Yaşam tablosuna geçmeden önce çalışmada izlem sürelerine uygun olacak şekilde izlem sürelerini sınıflara bölmek gerekir. Örneğin hastaların ortalaması 12 ay süreyle izlendiği bir çalışmada 1 ya da 2 aylık sınıflar uygun olabilir. Öte yandan ortalaması izlem süresi 6 aysa, 2 ya da 4 haftalık sınıflar daha uygun olacaktır.

Örnek 73'teki çalışmada izlem süresinin birer aylık sınıflara bölündüğünü düşünelim. Bu durumda yaşam tablosunun ilk sütununda 0-1 ay, 1-2 ay, ..., olarak izlem süresi aralıkları bulunacaktır (Tablo 3). Androjen grubu için yaşam tablosunu yaptığımızda 0-1 ay zaman aralığının başlangıcı olan 0. ayda 30 hasta sağ olarak çalışmaya alınmış olduğu için ikinci sütunun ilk sırasına 30 rakamı girelecektir.

Tablo 3 Örnek 73'teki çalışmada androjen grubunun yaşam tablosu.

İzlem süresi aralığı (ay)	Zaman aralığı başında sağ olanlar	Zaman aralığı içinde ölenler	Zaman aralığı içinde sağ olarak çalışmadan çıkanlar	Ölüm olasılığı	Sağkalım olasılığı	Kümülatif sağkalım olasılığı
n _i	d _i	w _i	q _i	p _i	S _i	
0-1	30	2	1	2/29.5 = 0.07	0.93	0.93
1-2	27	2	1	2/26.5 = 0.08	0.92	0.86
2-3	24	0	1	0/23.5 = 0.00	1.00	0.86
3-4	23	0	0	0/23 = 0.00	1.00	0.86
4-5	23	1	0	1/23 = 0.04	0.96	0.83
5-6	22	3	1	3/21.5 = 0.14	0.86	0.71
6-7	18	1	0	1/18 = 0.06	0.94	0.67
7-8	17	2	2	2/16 = 0.12	0.88	0.59
8-9	13	0	1	0/12.5 = 0.00	1.00	0.59
9-10	12	4	2	4/11 = 0.04	0.96	0.57
10-11	6	0	0	0/6 = 0.00	1.00	0.57
11-12	6	1	1	1/5.5 = 0.18	0.82	0.47
12-13	4	3	1	3/3.5 = 0.86	0.14	0.07

Çalışmaya alındıktan sonra bir aydan daha kısa sürede ölen hasta sayısı iki ise, üçüncü sütununa "2", ve bir aydan daha kısa bir süre izlendikten sonra kontrole gelmeyen ya da analize başlandığında izlem süresi bir aydan kısın olanların sayısı bir ise son sütuna "1" girilecektir.

Özetle çalışmaya alınan 30 hastadan, bir ay sonra geriye 27 hasta kalmış olacaktır. Başka bir deyişle, 1-2 ay zaman aralığının başlangıcında 27 hasta bulunacaktır. Bu 27 hastadan ikisinin 2 aydan daha kısa sürede öldüğü ve birinin de izlem süresinin 2 aydan kısa olduğu biliniyorsa, ikinci sıraya 2 ve 1 yazılacaktır. İlk ay bittiğinde 27 hastadan geriye 24 hasta kalmıştır. Bu şekilde hasta sayısı 0'a inene kadar işleme devam edilerek tablo doldurulur.

Yaşam tablosunda n_i , i . ay başlangıcındaki hasta sayısını, d_i , i . ay içinde ölen (*die*) hasta sayısını ve w_i ise i . ay içinde sağ olarak çalışmadan çıkan (*withdrawn*) hasta sayısını göstermektedir.

Bu tablodaki verilerden yararlanarak her zaman aralığı için ölüm olasılığı (q_i), sağkalım olasılığı (p_i) ve kümülatif sağkalım olasılığı (S_i) hesaplanabilir. İlk zaman aralığında 30 hastadan ikisi öldüğüne göre, ölüm olasılığı $2/30'$ dur, ama bu durumda bu süre boyunca izlenmiş ve en son görüldüğünde sağ olan hastaların verileri kullanılmamış olacaktır. Buna göre düzeltildiğinde ölüm olasılığı aşağıdaki formüle göre hesaplanır.

$$q_i = \frac{d_i}{n_i - w_i/2}$$

Sağkalım olasılığı p_i , $1-q_i$ 'ye eşittir. q_i ve p_i değerleri, yalnızca ilgili zaman aralığına ait olasılıklardır. Örneğin 5-6 ay zaman aralığındaki ölüm olasılığı olan %14, beşinci aya kadar sağ kalmış olan kişilerin, beşinci ay içinde ölmeye olasılığıdır. Aynı şekilde tedaviye başladıkten sonra beş ay yaşamış bir hastanın en az bir ay daha yaşama şansı %86'dır.

Birinci ay bittiğinde hastaların sağkalım olasılığı %93 olarak hesaplanmıştır. İkinci aya sağ olarak giren hastaların ikinci ay bittiğinde sağkalım olasılığı ise %92 olarak hesaplanmıştır. O halde çalışmaya alınan bir hastanın iki ay bittiğinde sağkalım olasılığı bu iki oranın çarpımı ($0.93 \times 0.92 = 0.86$) yani %86 olacaktır. Bu rakam, iki aylık kümülatif sağkalım olasılığıdır. Tüm zaman aralıkları için kümülatif sağkalım olasılığı aşağıdaki şekilde hesaplanabilir.

$$S_i = p_1 \cdot p_{(i-1)} \cdot p_{(i-2)} \cdot \dots \cdot p_2 \cdot p_1$$

Genellikle sağkalım analizi sonuçları tablo ile değil, sağkalım grafikleri ile verilir. Bu grafiklerde sağkalım eğrisinin güven sınırları da çizilebilir.

Yaşam tablosu yönteminde belirli bir zaman aralığında sağ olarak çalışmadan çıkan hastaların, zaman aralığının ortasında çalışmadan çıktıığı varsayıılır. Örneğin 0-1 ay zaman aralığında çalışmadan çıktıığı kişinin, bu zaman aralığının ortasında, yani çalışmaya alındıktan iki hafta sonra çalışmadan çıktıığı varsayıılır. Bu nedenle, eğer zaman aralıkları geniş tutulursa, ya da bu şekilde çalışmadan çıktı hasta sayısı çok fazlaysa, hesaplamalarda hatalar olabilir. Kaplan-Meier yönteminde bu sorun yoktur.

Yaşam tablosu analizinin varsayımları

Yaşam tablosu yönteminin güvenilir olabilmesi için bazı koşullara uygunluğun sağlanması gereklidir. Bunlardan biri, izlem süresi boyunca hastalara sağlanan koşulların değişmemesidir. Örneğin 1975-1992 yılları arasında izlenen meme tümörlerinin radyoterapi sonrası sağkalım analizlerini yapmak için bu 17 yıl boyunca hastalara sağlanan koşulların değişmemesi gereklidir. Eğer 1982 yılında yeni bir radyoterapi cihazı gelmiş, 1985 yılında güçlü bir kemoterapi protokoli tedaviye eklenmişse ve 1991 yılından itibaren de yüksek doz kemoterapi + otolog kemik iliği transplantasyonu uygulanmışsa, sağkalım analizini bu değişiklikleri göz ardı ederek yapmak, kesinlikle hatalıdır.

Yaşam analizi yapılrıken, çalışma bittiğinde sağ kalan ya da son kontrole sağ olarak görüldükten sonra kontrole gelmeyen hastaların da yarısının "riske maruz kalan hasta sayısı"na eklendiğini gördünüz. Örneğin 200 meme kanserinin radyoterapi sonrası sağkalım analizi yapılrıken, 45 hastanın bir süre sonra kontrole gelmediğini varsayıyalım. Bu hastaların yarısı, son izleni sürelerine karşılık gelen zaman aralığındaki "riske maruz kalan hastaların sayısı"na eklenecektir.

Bu hastalar son kez görüldükten sonra öldükleri için gelmemişlerse, yani bu grupta ölüm oranı yüksekse, hesapladığımız sağkalım oranları gerçeğin biraz üstünde olacaktır. Aksine, hastalar tamamen iyileşmiş ve bu nedenle kontrollere gelmek gereğini duymamışlarsa, hesapladığımız sağkalım oranları gerçeğin altında olacaktır. İşte sağkalım analizindeki varsayımlardan biri de çalışmadan sağ olarak çıktı hastaların sağkalım oranlarının, tüm çalışma grubundan farklı olmadığıdır.

KAPLAN-MEIER SAĞKALIM ANALİZİ

Bu yöntemde yaşam tablosuna göre iki farklılık vardır. Birincisi, izlem süresinin belirli zaman aralıklarına bölünmemesi, ikincisi ise ölüm olasılığı (α)

hesaplanırken, sağ olarak çalışmadan çıkanların hesaplamaya dahil edilmesidir. Yani $q_i = d_i / n_i$ 'dir.

Kaplan-Meier yönteminde önce yaşam tablosundaki zaman aralıkları yine, ilgilenilen olayın (*event*) gerçekleştiği, yani ölüm olan zamanlar ilk sütuna yazılır. Diğer hesaplamalar ölüm olasılığı hariç yaşam tablosunda olduğu gibidir. Sağkalım olasılığı ve kümülatif sağkalım olasılığı da, ölüm olasılığına bağlı olduğu için az da olsa farklı sonuçlar elde edilir.

Kaplan-Meier yönteminde, yaşam tablosu yönteminde olduğu gibi sağkalım süreleri sınıflandırılmadığı için, yaklaşık değil tam sağkalım olasılıkları elde edilir.

SAĞKALIM ANALİZİNDE RİSK FONKSİYONU

Risk fonksiyonu (*hazard function*), belirli bir zamana kadar sağ kalmış olan bir kişinin bir sonraki zaman aralığında ölüm olasılığına verilen addır. *Hazard* sözcüğünün ilk harfi olan H ile gösterilen risk fonksiyonu ya da risk hızı aşağıdaki formüle hesaplanır. Bu formülde d toplam ölüm sayısını, Σf ölümler olana kadar geçen toplam süreyi ve Σc ise sansürülü gözlemlerin toplam süresini (yani sağ kalan hastaların izlem sürelerinin toplamı) temsil etmektedir.

$$H = \frac{d}{\Sigma f + \Sigma c}$$

Aplastik anemi örneğinde 30 hastadan 19'unun olduğunu ve bu hastaların sağkalım sürelerinin toplamının 41 ay olduğunu varsayıyalım. Geriye kalan 11 sansürülü gözlemin izlem süreleri toplamı ise 76 ay olsun. Bu durumda

$$H = \frac{19}{41 + 76} = \frac{19}{117} = 0.16$$

Yani bir aylık ölüm hızı %16'dır. Eğer hasta sayısı fazlaysa, H değerinin tersi, yaklaşık olarak ortalama sağkalım süresini verir. Bu örnekte ortalama sağkalım süresi 6.1 ay (117/19) olarak hesaplanabilir.

$$\text{Ortalama sağkalım süresi} = \frac{1}{H}$$

Ortalama sağkalım süresinin %95 güven sınırları aşağıdaki formüle göre hesaplanabilir.

$$\%95 \text{ güven sınırları} = X \pm 1.96$$



Aplastik anemi çalışmasında ortalama sağkalım süresi ve %95 güven sınırları 6.1 ± 2.7 ay olarak bulunmuştur.

İKİ SAĞKALIM EĞRİSİNİN KARŞILAŞTIRILMASI

Yukarıdaki örnekte androjen grubundaki aplastik anemilerin sağkalım analizi ile ilgilendik. Çalışmada bir de immunosupresif tedavi grubu yer almıştır. Bu grup içinde aynı şekilde yaşam tablosu ya da Kaplan-Meier yöntemi ile sağkalım oranları hesaplanıp, sağkalım eğrisi çizilebilir. Peki iki tedavi grubu arasında sağkalım açısından fark var mıdır? Tedavilerden biri, diğerinden üstün müdür?

Sağkalım oranları açısından yalnızca farklı tedavi grupları değil, прогнозa etki eden etkenler açısından farklı hasta gruplarının da karşılaştırılması sıkılıkla gereklidir. Örneğin tümör evresi açısından dört farklı gruba ayrılan hastaların sağkalım oranlarının farklı olup olmadığı bilinmesi çok yararlıdır. Eğer tüm hastalar çalışma bitmeden önce ölmüş olsalardı, izlem süreleri sansürsüz olacağı için iki grup için Mann-Whitney U testi, ikiden çok grup için Kruskal-Wallis testi ile bu sorun çözülebilirdi. (İzlem süreleri hemen hemen her zaman logaritmik dağılım gösterdikleri için nonparametrik testler örnek olarak verilmiştir). Ancak izlem sürelerinin bir kısmı sansürül olacağı için bu yöntemler uygun değildir. Bu nedenle sansürül değişkenlerin karşılaştırılması için yöntemler geliştirilmiştir. Bunların başında zahmetli tablolar ve uzun hesaplamalar olduğu için elle yapılması zaman alıcıdır.

Gehan (genelleştirilmiş Wilcoxon) testi

Gehan testi, Wilcoxon işaret testinin sansürül gözlemler için geliştirilmiş şeklidir. Breslow testi ya da ikiden çok grup olduğunda genelleştirilmiş Kruskal-Wallis testi olarak da adlandırılır.

Bu yöntemde iki grupta yer alan hastalar ikişer ikişer birbirleriyle karşılaştırılır. Aplastik anemi çalışmasında iki grupta da 30'ar hasta olduğu toplam ikili karşılaştırma sayısı $30 \times 30 = 900$ olacaktır. Örneğin androjen grubu

bundaki birinci hasta ile immünosupresif grubundaki birinci hastanın sağkalım süreleri karşılaştırıldığında, eğer birinci gruptaki hastanın yaşam süresi, diğerinden uzunsa, +1, kısaysa -1 olarak kodlanır. Sansürlü gözlemler nedeniyle hangisinin sağkalım süresinin daha uzun olduğu belirlenmiyorsa, 0 olarak kodlanır. 900 ikili hasta eşi için bu işlem yapıldıktan sonra hesaplanan V* değerine Gehan test istatistiği adı verilir. Z tablosundan V* değerine karşılık gelen p değeri $p < 0.05$ ise iki sağkalım eğrisinin birbirinden farklı olduğu sonucuna varılır.

Logrank testi

İki sağkalım eğrisinin karşılaştırılmasında sık olarak kullanılan diğer bir test ise logrank testidir. Literatürde Mantel logrank istatistiği, Cox-Mantel logrank istatistiği ya da kısaca logrank istatistiği olarak adlandırılmıştır. Peto-Peto logrank testi ile de aynı sonuca ulaşılır.

Logrank testinin mantığı, iki oranın karşılaştırıldığı ki-kare testinin mantığına benzer. Her bir zaman aralığı için gruppardaki gözlenen ölüm sayıları, beklenen ölüm sayıları ile karşılaştırılarak hesaplanan ki-kare değeri, eğer tablo ki-kare değerinden büyükse (bu durumda $p < 0.05$ olacaktır), iki sağkalım eğrisinin birbirinden farklı olduğu sonucuna varılır. Ancak iki sağkalım eğrisinin ne kadarlık bir izlem süresi sonunda birbirinden ayrıldığını logrank testi belirtmez.

İkiden fazla grup karşılaştırılıyorsa, ikişerli karşılaştırmaların da yapılması gereklidir. Bu durumda karşılaştırılan iki grubun sağkalım eğrilerinin farklı olduğu sonucuna varmak için yanlışma payını 0.05'den daha aşağıya çekmek gerekdir. Örneğin tümör evresine göre 4 farklı gruba ayrılan hastaların sağkalım eğrilerinin logrank testi ile farklı olarak saptandığını varsayılmı. Evre I-II, evre II-III ve evre III-IV arasında sağkalım eğrilerinin farklı olup olmadığı test edilirken, üç ayrı ikişerli karşılaştırma yapıldığı için $p < 0.05$ yerine $p < 0.017$ ($0.05 / 3$) ise iki grup arasında fark olduğu söylenebilir. P değerleri Evre I-II için 0.13, evre II-III için 0.03 ve evre III-IV için 0.001 ise, yalnızca evre III ile evre IV için sağkalım eğrilerinin farklı olduğu söylenebilir. Evre II ile III'ün karşılaştırılmasında $p > 0.017$ olduğu için bu iki grup arasında fark olduğu söylenenemez.

İkinci bir değişkene göre kategorize edilmiş altgrupların sağkalım eğrileri de karşılaştırılabilir. Örneğin tümör evrelerine göre sağkalım eğrileri farklı olabilir, ama erkek ve kadınlarda durumun aynı olması gerekmekz. Erkeklerde evre III'de, evre II'ye göre sağkalım süresi daha kısa iken, kadınlarda durum farklı olabilir. Bu nedenle evreye göre belirlenmiş dört farklı sağkalım eğrisinin analizinde cinsiyete göre tabakalandırma gerekebilir. Tabakalandırılmış

sağkalım analizi, sağkalım süresine etkisi araştırılan asıl değişkenin (örneğin tümör evresi) etkisinden tabakalandırıcı değişkenin (örneğin cinsiyetin) etkilerinin arıtılmasını sağlar. Tabakalandırılmış sağkalım analizinde kullanılan yöntemlerden biri olan Mantel-Haenszel ki-kare testi, logrank testinin genelleştirilmiş bir şeklidir. Logrank testine ek olarak birden çok kategorik değişkenin farklı düzeyleri arasında sağkalım eğrileri açısından fark olup olmadığı incelenebilir. Örneğin aplastik anemide iki farklı tedavinin sağkalım süresi üzerindeki etkileri erkek ve kadınlarda farklı ise, tedavi gruplarının sağkalım sürelerindeki farklılığın nedeninin cinsiyetle ilişkili olup olmadığı bu yöntemle saptanabilir. Bu anlamda, logrank testi, varyans analizine, Mantel-Haenszel ki-kare testi ise kovaryans analizine benzetilebilir.

İlk kez 50'li yıllarda tanımlanan Kaplan-Meier yönteminden sonraki yıllar içinde çeşitli zamanlarda tanımlanan sağkalım analizleri yaklaşık olarak aynı sonuçları vermektedir. Bu nedenle tıp literatüründe çeşitli adlarla belirtilen sağkalım analiz yöntemlerinin hangisinin kullanıldığına pek ömeli yoktur. Yalnız logrank grubu testler, Gehan testinden anlamı ve uygulanabilirliği açısından ayırlır.

Gehan testinde testin sonucunu erken dönemdeki ölümler geç dönemdeki ölümlere göre daha fazla etkilerken, logrank testinde tüm zamanlar için ölümlerin ortaya çıkışının eşit ağırlıktadır. Öte yandan, logrank testinde farklı gruptardaki hastaların risk hızları (*hazard rate*) oranının, tüm zamanlarda aynı olduğu varsayılar. Örneğin androjen tedavisi verilenlerde erken dönem ölümleri azalırken, immünosupresif tedavi verilenlerde geç dönem ölümleri azalıyorsa, bu koşul sağlanamadığı için logrank testi yerine Gehan testi uygulanmalıdır. Risk hızlarının oransal olduğu durumlarda Cox oransal risk modeli (*Cox proportional hazard model*) uygulanabilir.

COX REGRESYON ANALİZİ

İki farklı tedavinin sağkalım süresine etkisini incelerken, farklı tedavilerin verildiği hastaların sağkalım süresini etkileyen diğer faktörler açısından benzer olması sağlanmalıdır. Aksi takdirde bir grupta sağkalım süresinin diğer gruptan farklı olması durumunda bunun tek nedeninin tedavinin etkisi olduğunu öne sürmek mümkün olmaz. Örnek 73'teki çalışmada hangi tedavi verilirse verilsin sağkalım süresi yaşıtlarda daha kısa ise ve tedavi gruppından biri diğerinden daha yaşlı ise sağkalım süreleri arasındaki farka yaşın etkisinin de dikkate alınması gereklidir. Yaşam tablosu ya da Kaplan-Meier sağkalım analizinde gruplayıcı değişken kategorik olmak zorundadır, bu nedenle sayısal bir değişken olan yaşın da bu analiz modellerine dahil edilmeli

mümkün değildir. Bu örnekte olduğu gibi sağkalım süresi vb. sansürlü değişkenler üzerine etkileri incelenen değişkenler arasında sayısal değişkenler de varsa Cox regresyon analizi uygulanabilir.

SPSS'TE SAĞKALIM ANALİZİ NASIL YAPILIR?

Yaşam tablosu

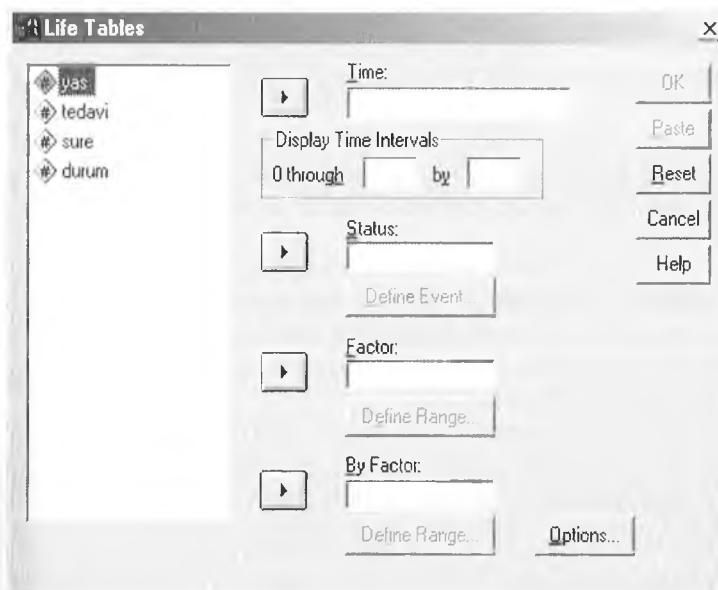
Analyze >> Survival >> Life Tables

Yukarıdaki seçenekle Life Tables diyalog kutusu açılır (Şekil 1).

Time penceresine izlem süresini seçin.

Display time intervals penceresine zaman aralıklarını 0'dan başlamak üzere yazılacaktır. Bu pencereye maksimum izlem süresinden daha uzun bir süreyi ve istediğiniz zaman aralığını girin. Örneğin maksimum izlem süresinin 55 ay olduğu bir çalışmada bu pencereye 0 through 60 by 3 girdiğinizde 3'er aylık zaman aralıkları ile 60. aya kadar yaşam tablosu yapılacaktır.

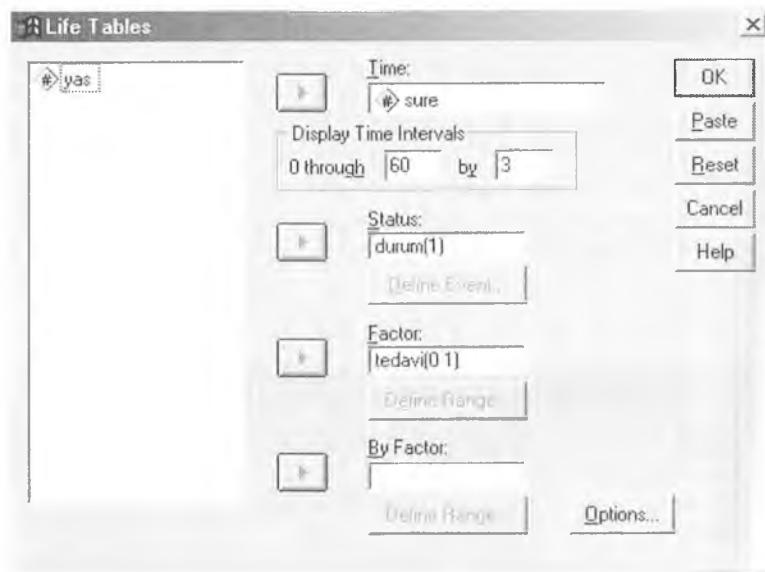
Status penceresine hastanın son izlemindeki durumunu gösteren değişkeni seçin. Örneğin durum adlı bir değişkene hasta öldüyse "1", sağsa "0" olarak girilmişse, Status penceresine durum değişkenini seçin. Daha sonra Define Event'ten "oldu"ye "1"in karşılık geldiğini belirleyin.



Şekil 1
SPSS'te "Life Tables" diyalog kutusu.

Factor penceresine bir değişken seçilmemişse, veritabanındaki tüm hastaların yaşam tablosu yapılır. Altgrupların ayrı ayrı yaşam tablolarını elde etmek için *Factor* penceresine altgrupları belirten değişkeni seçin. Örneğin tedavi adlı değişken, androjen verilen hastalar için "0", immüñosupresif verilen hastalar için "1" olarak kodlanmışsa, *Factor* penceresine tedavi değişkenini seçin ve *Define Range*'den minimumu "0", maksimumu "1" olarak belirleyin.

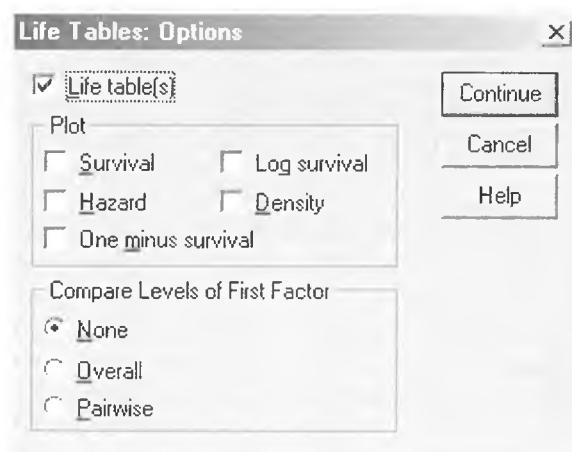
Örnek 73'teki çalışmada androjen ve immüñosupresif tedavi gruplarının yaşam tablosu analizi için **Life Tables** diyalog kutusu Şekil 2'deki gibi olmalıdır.



Şekil 2
SPSS'te
"Life
Tables"
diyalog
kutusu.

By Factor penceresine ikinci bir gruplayıcı değişken daha girebilirsiniz. Örneğin buraya cinsiyet değişkenini seçerseniz, erkek ve kadınlar için ayrı ayrı olmak üzere androjen ve immüñosupresif grubunun yaşam tabloları, ya ni dört adet yaşam tablosu elde edersiniz.

Options seçeneği ile açılan pencereki *Plot* seçeneğinden sağkalım (*survival*) ve ölüm olasılığı (*hazard*) eğrileri çizilebilir. *Factor* penceresine girilmiş değişkenin belirlediği altgrupların yaşam eğrilerinin Gehan testi ile karşılaştırılması global olarak (*overall*), ya da ikişerli olarak (*pairwise*) bu pencereden çilebilir (Şekil 3).



Şekil 3
SPSS'te "Life Tables" diyalog kutusundaki "Options" menüsü.

Şekil 2'deki seçim yapıldıktan sonra OK'lendiğinde Şekil 4'teki yaşam tablosu çıkacaktır. Bu tablodaki rakamların ne anlamına geldiğini bir örnekle inceleyelim. Görüldüğü gibi sağkalımla ilgili veriler her üç aylık zaman aralığı için verilmiştir.

Örneğin dikkatimizi androjen grubunda 15-18. aylar arasındaki olaylara odaklayalım. Bu zaman aralığına ilişkin rakamlar Interval start time sütununda "15" yazan satırda belirtilmiştir. Onbeşinci ayda 22 hastanın bulunduğu, 15-18. aylar içinde bir hastanın çalışmadan sağ olarak çıktığı ve bir hastanın da öldüğü görülüyor. Bunlara dayanarak 15-18. ay zaman aralığında ölüm oranının %4.65, sağkalım oranının %95.35 ve kümülatif sağkalım oranının %84.81 olduğu görülmüyör. SPSS'ten elde edilen yaşam tablosundaki terimler ve bunların ne anlamına geldikleri Tablo 4'te belirtilmiştir.

Options menüsündeki *Plots*'tan *Survival* ve *Compare Levels of First Factor*'den *Pairwise* seçildiğinde iki tedavi grubunun sağkalım eğrileri çizilecek ve iki grubun sağkalım oranları arasındaki farkın istatistiksel olarak önemli olup olmadığı test edilecektir (Şekil 5 ve 6).

Life Table

	Survival Variable		SURE		izlem süresi (ay)				
	for	TEDAVI	Tedavi	=	0	Androjen			
Interval Start Times	Number Entrg this	Number Wdrawn During	Number Exposd to Risk	Number of Termnl Events	Propn Termi-nating	Propn Sur-viving	Cumul Propn Surv at End	Probabi-lity Densty	Hazard Rate
.0	30.0	2.0	29.0	.0	.0000	1.0000	1.0000	.0000	.0000
3.0	28.0	1.0	27.5	1.0	.0364	.9636	.9636	.0121	.0123
6.0	26.0	.0	26.0	1.0	.0385	.9615	.9266	.0124	.0131
9.0	25.0	.0	25.0	1.0	.0400	.9600	.8895	.0124	.0136
12.0	24.0	2.0	23.0	.0	.0000	1.0000	.8895	.0000	.0000
15.0	22.0	1.0	21.5	1.0	.0465	.9535	.8481	.0138	.0159
18.0	20.0	.0	20.0	1.0	.0500	.9500	.8057	.0141	.0171
21.0	19.0	.0	19.0	.0	.0000	1.0000	.8057	.0000	.0000
24.0	19.0	1.0	18.5	1.0	.0541	.9459	.7622	.0145	.0185
27.0	17.0	.0	17.0	1.0	.0588	.9412	.7173	.0149	.0202
30.0	16.0	1.0	15.5	.0	.0000	1.0000	.7173	.0000	.0000
33.0	15.0	.0	15.0	1.0	.0667	.9333	.6695	.0159	.0230
36.0	14.0	2.0	13.0	2.0	.1538	.8462	.5665	.0343	.0556
39.0	10.0	.0	10.0	1.0	.1000	.9000	.5099	.0189	.0351
42.0	9.0	1.0	8.5	1.0	.1176	.8824	.4499	.0200	.0417
45.0	7.0	.0	7.0	.0	.0000	1.0000	.4499	.0000	.0000
48.0	7.0	1.0	6.5	3.0	.4615	.5385	.2422	.0692	.2000
51.0	3.0	.0	3.0	3.0	1.0000	.0000	.0000	.0807	.6667

The median survival time for these data is 42.49

Life Table

	Survival Variable		SURE		izlem süresi (ay)				
	for	TEDAVI	Tedavi	=	1	IS tedavi			
Interval Start Time	Number Entrg this	Number Wdrawn During	Number Exposd to Risk	Number of Termnl Events	Propn Termi-nating	Propn Sur-viving	Cumul Propn Surv at End	Probabi-lity Densty	Hazard Rate
.0	30.0	1.0	29.5	.0	.0000	1.0000	1.0000	.0000	.0000
3.0	29.0	3.0	27.5	.0	.0000	1.0000	1.0000	.0000	.0000
6.0	26.0	.0	26.0	1.0	.0385	.9615	.9615	.0128	.0131
9.0	25.0	1.0	24.5	1.0	.0408	.9592	.9223	.0131	.0139
12.0	23.0	.0	23.0	1.0	.0435	.9565	.8822	.0134	.0148
15.0	22.0	2.0	21.0	.0	.0000	1.0000	.8822	.0000	.0000
18.0	20.0	.0	20.0	1.0	.0500	.9500	.8381	.0147	.0171
21.0	19.0	.0	19.0	1.0	.0526	.9474	.7940	.0147	.0180
24.0	18.0	1.0	17.5	.0	.0000	1.0000	.7940	.0000	.0000
27.0	17.0	.0	17.0	1.0	.0588	.9412	.7473	.0156	.0202
30.0	16.0	1.0	15.5	.0	.0000	1.0000	.7473	.0000	.0000
33.0	15.0	1.0	14.5	.0	.0000	1.0000	.7473	.0000	.0000
36.0	14.0	2.0	13.0	2.0	.1538	.8462	.6323	.0383	.0556
39.0	10.0	1.0	9.5	.0	.0000	1.0000	.6323	.0000	.0000
42.0	9.0	1.0	8.5	.0	.0000	1.0000	.6323	.0000	.0000
45.0	8.0	1.0	7.5	.0	.0000	1.0000	.6323	.0000	.0000
48.0	7.0	1.0	6.5	1.0	.1538	.8462	.5350	.0324	.0556
51.0	5.0	3.0	3.5	.0	.0000	1.0000	.5350	.0000	.0000
54.0	2.0	.0	2.0	2.0	1.0000	.0000	.0000	.1783	.6667

The median survival time for these data is 54.20

Şekil 4 Örnek 73'teki çalışmada her iki gruptaki hastalara ilişkin yaşam tablosu analizi. Üstte androjen grubu, altta immüenosupresif tedavi grubu sonuçları görülmektedir.

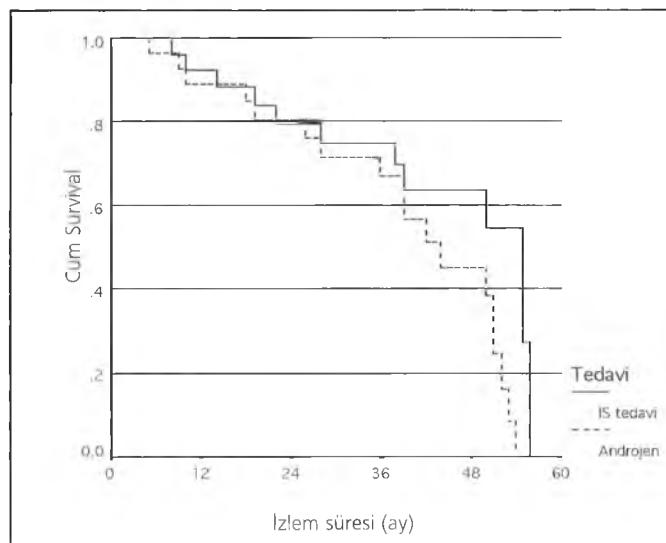
Tablo 4 SPSS'te Yaşam Tablosu Analizi'nde kullanılan terimler ve açıklamalar.

TANIM	AÇIKLAMA
<i>Interval Start Time</i>	Zaman aralığının başlangıç zamanı (Bitiş zamanı ise bir sonraki zaman aralığının başlangıcıdır)
<i>Number Entering This Interval</i>	n_i Zaman aralığının başlangıcında sağ olan hasta sayısı
<i>Number Withdrawn During This Interval</i>	w_i Zaman aralığının başlangıcında sağ olan ve zaman aralığı içinde çalışmadan sağ olarak çıkan hasta sayısı ¹
<i>Number Exposed to Risk</i>	Zaman aralığı içinde riske maruz kalan hasta sayısı ²
<i>Number of Terminal Events</i>	d_i Zaman aralığının içinde ölen hasta sayısı
<i>Proportion of Terminal Events</i>	q_i Ölüm oranı ya da olasılığı ³
<i>Proportion Surviving</i>	p_i Sağkalım oranı ya da olasılığı ⁴
<i>Cumulative Proportion</i>	
<i>Surviving at End</i>	S_i Kumülatif sağkalım oranı ya da olasılığı ⁵
<i>Probability Density</i>	Birim zamanda ölüm olayınin gerçekleşmesi olasılığı
<i>Hazard Rate</i>	H Risk hızı ⁶
<i>Standard Error of the Cumulative</i>	
<i>Proportion Surviving at End</i>	
<i>Standard Error of the Probability Density</i>	Yukarıdaki üç parametrenin standart hataları
<i>Standard Error of the Hazard Rate</i>	
<i>Median Survival Time</i>	Ortanca sağkalım süresi ⁷

1. Bu hastalar son kontrolde görüldüklerinde henüz sağdır; ya çalışma bittiğinde sağdırılar ya da son kontrolde sağ olarak görüldükten sonra kontrole gelmemiştir. Bu hastalara ait izlem süreleri sansürlüdür.
2. Bu sayı zaman aralığına sağ olarak giren hasta sayısından sağ olarak çalışmadan çıkan hasta sayısının yarısının çıkarılması ile hesaplanır. Burada sağ olarak çalışmadan çıkan hastaların yarısı yerine, hiçbir ya da hepsi de hesaplamaya dahil edilebilir. Bunu belirleyen durum, bu hastaların neden kontrole gelmediğidir.
3. Ölen hasta sayısının riske maruz kalan hasta sayısına oranı.
4. Sağkalım oranı = 1 - Ölüm oranı.
5. İlgili zaman aralığı ve daha önceki zaman aralıklarındaki sağkalım oranlarının çarpımı olarak hesaplanır.
6. İlgili zaman aralığına sağ olarak giren bir hastanın zaman aralığı içinde ölmeye olasılığı.
7. Kumülatif sağkalım olasılığının 0.50 olduğu zamana karşılık gelen süredir. Eğer son zaman aralığındaki kümülatif sağkalım oranı hala 0.50'den büyükse, ortanca sağkalım süresi "+" işaretli ile gösterilir.

Şekil 5'deki iki tedavi grubunun sağkalım eğrileri arasında fark olmadığı görülmektedir. Şekil 6'da androjen grubundaki hastalardan 12'si (%40), imünosupresif tedavi grubundaki hastalardan 19'unun (%63.33) izlem sürelerinin sansürlü olduğu, yani halen yaşadıkları anlaşılmaktadır. Prob ile gösterilen p değeri 0.35'dir; yani iki tedavi grubundaki sağkalım oranları arasındaki fark istatistiksel olarak önemli değildir.

Şekil 5
Örnek 73'teki çalışmada grupların yaşam tablosu yöntemine göre elde edilmiş sağkalım eğrileri. Yatay eksende ay olarak süre, dikey eksen ise sağkalım olasılığı bulunmaktadır. Dikey eksenin 0.2 %20'ye, 1.0 ise %100'e karşılık gelmektedir.



Comparison of survival experience using the Wilcoxon (Gehan) statistic

Survival Variable grouped by	SURE TEDAVI	İzlem süresi (ay) Tedavi
Overall comparison statistic	.866	D.F. 1 Prob. .3521
Group label	Total N	Uncen Cen Pct Cen Mean Score
0 Androjen	30	18 12 40.00 -2.7000
1 IS tedavi	30	11 19 63.33 2.7000

Şekil 6 Örnek 73'teki çalışmada iki grubun sağkalım oranlarının Gehan testi ile karşılaştırılmıştır.

Kaplan-Meier sağkalım analizi

Analyze >> Survival >> Kaplan-Meier

Yukarıdaki seçenekle Kaplan-Meier diyalog kutusu açılır (Şekil 7).

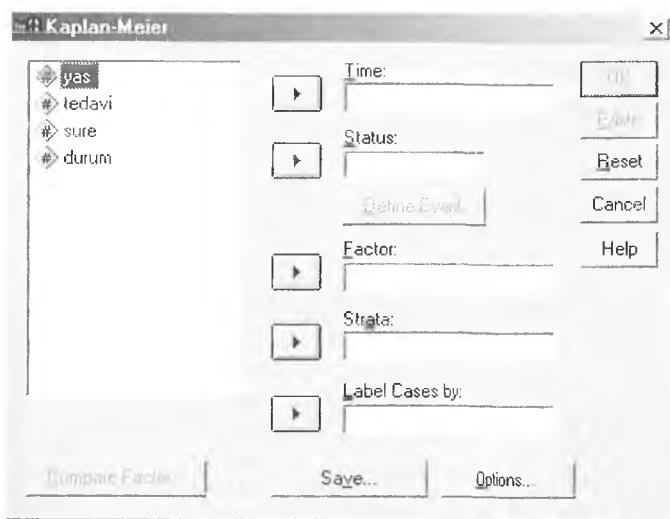
Time penceresine izlem süresini seçin.

Status penceresine hastanın son izlemindeki durumunu gösteren değişkeni seçin. Örneğin durum adlı bir değişkene hasta öldüyse "1", sağsa "0" olarak girilmişse, *Status* penceresine durum değişkenini seçin. Daha sonra *Define Event*'ten "öldü"ye "1"in karşılık geldiğini belirleyin.

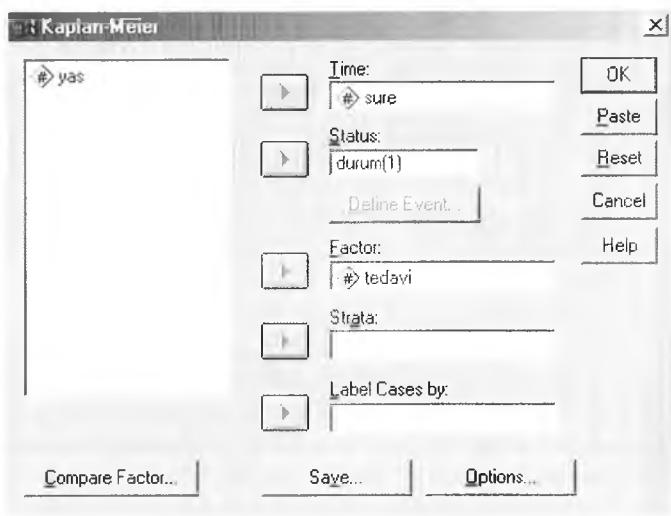
Factor penceresine bir değişken seçilmemişse, veritabanındaki tüm hastaların sağkalım analizi yapılır. Altgrupların ayrı ayrı sağkalım analizlerini elde etmek için *Factor* penceresine altgrupları belirten değişkeni seçin. Örneğin hastalara verilen tedavi, tedavi adlı değişkene kodlanarak girilmişse, *Factor* penceresine tedavi değişkenini seçin.

Strata penceresine tabakalandırıcı değişkeni girebilirsiniz. Örneğin buraya cinsiyet değişkenini seçerseniz, erkek ve kadınların sağkalım eğrilerini, androjen ve immüñosupresif grubunun sağkalım eğrileri ayrı ayrı olmak üzere aynı grafikte çizilmiş olarak elde edersiniz.

Örnek 73'teki çalışmanın analizini bu kez Kaplan-Meier yöntemi ile yapmak istersek, Kaplan-Meier diyalog kutusu Şekil 8'deki gibi olmalıdır.



Şekil 7
SPSS'te
"Kaplan-Meier"
diyalog kutusu.



Şekil 8
SPSS'te "Kaplan-Meier" diyalog kutusu.

Şekil 8'deki seçim yapıldığında çıktı Şekil 9'daki gibi olacaktır. Bu tabloda izlem süresi kısadan uzuna doğru hastalar dizilmiştir. *Time* sütununda izlem süresi, *Status* sütununda izlem süresi sonundaki durum yer almaktadır. Örneğin androjen grubunda bir hasta 1'inci ay sonunda, bir hasta 2'nci ay sonunda sağ olarak izlemden çıkışmışlardır, bir hasta 4'tüncü ayda ölmüştür, ..., bir hasta 53'üncü ayda ölmüştür. Bu grupta 12 hastanın (%40) yaşadığı, 18 ölüm gözleendiği, ortalama sağkalım süresinin 38 ay, ortanca sağkalım süresinin 43 ay olduğu anlaşılmaktadır.

Survival Analysis for SURE izlem süresi (ay)

Factor TEDAVI = Androjen

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
1	Yaşıyor			0	29
2	Yaşıyor			0	28
4	Öldü	.9643	.0351	1	27
5	Yaşıyor			1	26
8	Öldü	.9272	.0496	2	25
9	Öldü	.8901	.0599	3	24
12	Yaşıyor			3	23
13	Yaşıyor			3	22
16	Yaşıyor			3	21
17	Öldü	.8477	.0705	4	20
18	Öldü	.8053	.0787	5	19
24	Yaşıyor			5	18
25	Öldü	.7606	.0861	6	17
27	Öldü	.7159	.0919	7	16
3	Yaşıyor			7	15
35	Öldü	.6681	.0974	8	14
38	Öldü			9	13
38	Öldü	.5727	.1043	10	12
38	Yaşıyor			10	11
38	Yaşıyor			10	10
41	Öldü	.5154	.1084	11	9
43	Öldü	.4581	.1105	12	8
43	Yaşıyor			12	7
49	Öldü	.3927	.1124	13	6
50	Öldü			14	5
50	Öldü	.2618	.1064	15	4
50	Yaşıyor			15	3
51	Öldü	.1745	.1006	16	2
52	Öldü	.0873	.0796	17	1
53	Öldü	.0000	.0000	18	0

Number of Cases: 30 Censored: 12 (40.0%) Events: 18
 Survival Time Standard Error 95% Confidence Interval
 Mean: 38 3 (32, 44)
 Median: 43 7 (30, 56)

Factor TEDAVI = IS tedavi

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
2	Yaşıyor			0	29
4	Yaşıyor			0	28
4	Yaşıyor			0	27
5	Yaşıyor			0	26
7	Öldü	.9615	.0377	1	25
9	Öldü	.9231	.0523	2	24
10	Yaşıyor			2	23
13	Öldü	.8829	.0636	3	22
15	Yaşıyor			3	21
17	Yaşıyor			3	20
18	Öldü	.8388	.0741	4	19
21	Öldü	.7946	.0823	5	18
25	Yaşıyor			5	17
27	Öldü	.7479	.0898	6	16
30	Yaşıyor			6	15
35	Yaşıyor			6	14
37	Öldü	.6945	.0980	7	13
38	Öldü	.6411	.1040	8	12
38	Yaşıyor			8	11
38	Yaşıyor			8	10
41	Yaşıyor			8	9
43	Yaşıyor			8	8
46	Yaşıyor			8	7
49	Öldü	.5495	.1230	9	6
50	Yaşıyor			9	5
5	Yaşıyor			9	4
52	Yaşıyor			9	3
53	Yaşıyor			9	2
54	Öldü	.2747	.2038	10	1
55	Öldü	.0000	.0000	11	0

Number of Cases: 30 Censored: 19 (63.33%) Events: 11
 Survival Time Standard Error 95% Confidence Interval
 Mean: 43 4 (36, 50)
 Median: 54 9 (37, 71)

Survival Analysis for SURE izlem süresi (ay)		Total	Number Events	Number Censored	Percent Censored
TEDAVI	Androjen	30	18	12	40.00
TEDAVI	IS tedavi	30	11	19	63.33
Overall		60	29	31	51.67

Şekil 9
 SPSS'te Kaplan-Meier sağkalım analizi sonuçları.

Compare factor seçeneği ile açılan pencereden sağkalım eğrilerinin karşılaşırıldığında kullanılan Log rank, Breslow ve Tarone-Ware testlerini seçebilirsiniz (Şekil 10). Eğer faktör değişkeni ordinalse *Linear trend for factor levels* seçeneği ile faktör altgruplarının sağkalım eğrilerinin artma ya da azalma eğilimi test edilebilir. Örneğin bir ilaçın 10, 20, 30 ve 40 mg/gün gibi eşit aralıklı dört farklı dozu ile sağkalım süresi arasında aynı yönde doğrusal bir ilişki varsa, yani doz arttıkça sağkalım süresi artıyorsa, bu testle saptanabilir.



Şekil 10
SPSS "Kaplan-Meier"
diyalog kutusundaki
"Compare Factor
Levels" menüsü.

Seçilmiş olan *Pooled over strata* seçeneği ile tüm altgruplar global olarak karşılaştırılır. Eğer tabakalandırıcı değişken seçilmişse, ona göre düzeltilmiş olarak grupların global karşılaştırılması yapılır. *Pairwise over strata* seçeneği ile ikişerli grup karşılaştırmaları yapılır. *Strata* değişkeni seçilmişse ona göre düzeltilmiş olarak test yapılır. *For each stratum* seçeneği ile her tabaka için faktör değişkeninin altgrupları arasında global karşılaştırma yapılır.

Faktör olarak tümör evresi, strata olarak cinsiyet seçilmişse, erkekler ve kadınlar için ayrı ayrı olmak üzere, dört evre global olarak karşılaştırılır. *Pairwise for each stratum* seçeneği global karşılaştırma yerine ikişerli karşılaştırmaların yapılması dışında *for each stratum* seçeneği ile aynıdır.

Üç farklı tedavi yönteminin yaşam süresine etkisinin araştırıldığı bir çalışmada cinsiyetler arasında tedaviye yanıt açısından fark olduğu varsayımlı ile sağkalım analizinde faktör değişkeni olarak tedavi, tabaka değişkeni olarak cinsiyetin seçildiği durumlarda istatistik sonuçlarının nasıl olacağını ve tabaka değişkeninin kullanılıp kullanılmamasının sonucu nasıl etkileyeceğini ni Şekil 11'de görelim.

"Pooled over strata" seçilmiş, tabaka değişkeni yok.

Sonuç: Üç tedavi arasında fark yok ($p=0.79$).

	Statistic	df	Significance
Log Rank	.46	2	.7932

"Pooled over strata" seçilmiş, tabaka değişkeni: Cinsiyet.

Sonuç: Üç tedavi arasında, cinsiyete göre düzeltildiğinde de fark yok ($p=0.87$).

	Statistic	df	Significance
Log Rank	.27	2	.8717

"Pairwise over strata" seçilmiş, tabaka değişkeni yok.

Sonuç: Üç tedavi arasında ikişerli fark yok.

Log Rank Statistic and (Significance)

Factor	1	2
2	.13	
	(.7177)	
3	.11	.36
	(.7422)	(.5463)

"Pairwise over strata" seçilmiş, tabaka değişkeni: Cinsiyet.

Sonuç: Üç tedavi arasında, cinsiyete göre düzeltildiğinde de ikişerli fark yok.

Log Rank Statistic and (Significance)

Factor	1	2
2	.02	
	(.8885)	
3	.18	.02
	(.6749)	(.8949)

"For each stratum" seçilmiş, tabaka değişkeni: Cinsiyet.

Sonuç: İki cinsiyette de üç tedavi arasında ikişerli fark yok.

Test Statistics for Equality of Survival Distributions for TEDAVI

For CINS = ERKEK

	Statistic	df	Significance
Log Rank	2.03	2	.3624

For CINS = KADIN

	Statistic	df	Significance
Log Rank	5.24	2	.0729

"Pairwise for each stratum" seçilmiş, tabaka değişkeni: Cinsiyet.

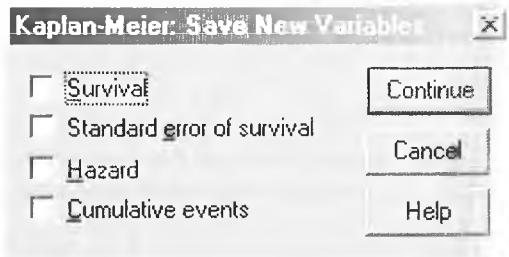
Sonuç: Erkeklerde üç tedavi arasında ikişerli fark yok. Kadınlarda 1 ve 2 kodlu tedaviler arasında fark olabilir ($p=0.0289$).

Log Rank Statistic and (Significance)

For CINS		= ERKEK	
Factor		1	2
2		1.95	
		(.1629)	
3		1.82	.02
		(.1775)	(.8898)
For CINS		= KADIN	
Factor		1	2
2		4.77	
		(.0289)	
3		3.46	.13
		(.0628)	(.7207)

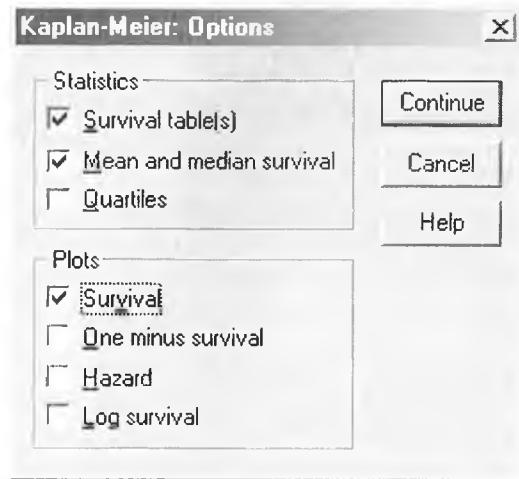
Şekil 11 Kaplan-Meier sağkalım analizinde grupların karşılaştırılmasında çeşitli seçeneklerle elde edilen sonuçlar (sayfa 271'deki tablolara da bakınız).

Save menüsü ile kümülatif sağkalım olasılıkları ve standart hataları (*survival*, *standard error of survival*) ve kümülatif risk olasılığı (*hazard*) dosyaya kaydedilebilir (Şekil 12).



Şekil 12
SPSS "Kaplan-Meier" diyalog kutusunda "Save" menüsü.

Options seçeneği ile açılan pencereden sağkalım tabloları ve ortalama ve ortanca sağkalım süreleri hesaplanabilir. Plot seçeneğinden sağkalım ve ölüm olasılığı (*hazard*) eğrileri çizilebilir (Şekil 13).



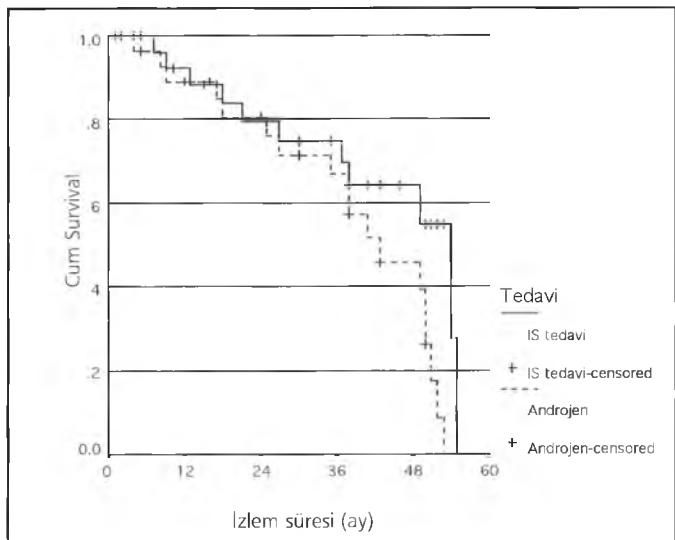
Şekil 13
SPSS "Kaplan-Meier" diyalog kutusunda "Options" menüsü.

Şekil 14'de Örnek 73'teki çalışmada grupların sağkalım eğrilerinin karşılaştırıldığı testlerin sonuçları ve Şekil 15'de ise sağkalım eğrileri görülmektedir. Şekil 14'deki üç testin sonucunda da *Significance* olarak verilen p değerinin 0.05'ten büyük olduğu, yani iki sağkalım eğrisi arasındaki farkın istatistiksel olarak önemli olmadığı anlaşılmaktadır.

Test Statistics for Equality of Survival Distributions for TEDAVI			
	Statistic	df	Significance
Log Rank	3.80	1	.0552
Breslow	.86	1	.3543
Tarone-Ware	1.74	1	.1872

Şekil 14 Örnek 73'teki çalışmada grupların sağkalım eğrilerinin karşılaştırıldığı testlerin sonuçları.

Şekil 15
 Örnek 73'teki
 çalışmada
 grupperin Kaplan-
 Meier yöntemine
 göre elde edilmiş
 sağkalım eğrileri.
 Yatay eksende ay
 olarak süre, dikey
 eksende ise
 sağkalım olasılığı
 bulunmaktadır.
 Dikey eksendeki 0.2
 %20'ye, 1.0 ise
 %100'e karşılık
 gelmektedir.



Cox regresyon analizi

Analyze >> Survival >> Cox Regression

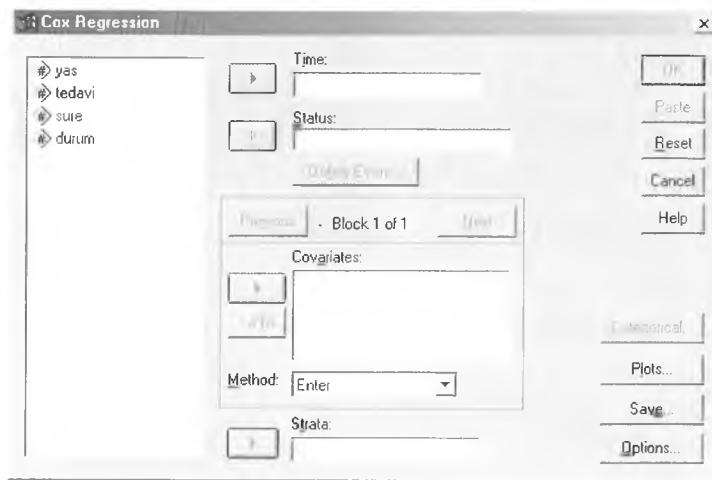
Yukarıdaki seçenekle **Cox Regression** diyalog kutusu açılır (Şekil 16).

Time penceresine izlem süresini seçin.

Status penceresine hastanın son izlemindeki durumunu gösteren değişkeni seçin. Örneğin durum adlı bir değişkene hasta öldüyse "1", sağsa "0" olarak girilmişse, *Status* penceresine durum değişkenini seçin. Daha sonra *Define Event*'ten "oldü"ye "1"in karşılık geldiğini belirleyin.

Covariates penceresine sağkalım süresine etkilerinin inceleneceği kategorik veya sayısal değişkenleri seçin. Örneğin hastalara verilen tedavi, tedavi adlı değişkene, hastaların yaşı yaşı adlı değişkene girilmişse, *Covariates* penceresine tedavi ve yaşı değişkenlerini seçin.

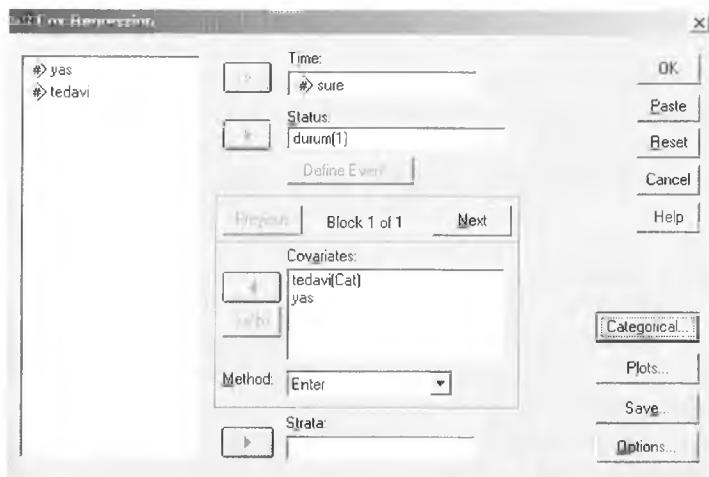
Categorical seçeneği ile açılan *Define Categorical Covariates* penceresinde tedavi değişkenini *Categorical Covariates* listesine seçin.



Şekil 16
SPSS'te "Cox Regression" diyalog kutusu.

Strata penceresine tabakalandırıcı değişkeni girebilirsiniz. Örneğin bura-ya cinsiyet değişkenini seçerseniz, erkek ve kadınların sağkalım eğrilerini, androjen ve immüñosupresif grubunun sağkalım eğrileri ayrı ayrı olmak üzere aynı grafikte çizilmiş olarak elde edersiniz.

Örnek 7'teki çalışmanın analizini bu kez yaşın etkisini de incelemeye alarak Cox regresyon analizi ile yapmak istersek, **Cox Regression** diyalog kutusu Şekil 17'deki gibi olmalıdır.



Şekil 17
SPSS'te "Cox Regression" diyalog kutusu.

Dependent Variable: SURE izlem süresi (ay)

	Chi-Square	df	Sig
Overall (score)	4.281	2	.1176

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
TEDAVI	-.7518	.4134	3.3068	1	.0690	-.0863	.4715
YAS	-.0112	.0158	.5035	1	.4780	.0000	.9888

Covariate Means

Variable	Mean
TEDAVI	.5088
YAS	38.2632

Şekil 18 SPSS'te Cox regresyon analizi sonuçları.

Şekil 17'deki seçim yapıldığında çıktı Şekil 18'deki gibi olacaktır. Bu tabloda sağkalım süresine etkilerine karşılık gelen p değerlerinin hem tedavi, hem de yaş için 0.05'ten büyük olduğu (tedavi için 0.0690, yaş için 0.4780) görülmektedir.

bölüm 16

ÇOK DEĞİŞKENLİ İSTATİSTİK ANALİZ YÖNTEMLER

Çok değişkenli analiz yöntemlerine giriş

Kovaryans analizi

Nominal değişkenlerin kestirimi

Lojistik regresyon

Diskriminant analiz

Lojistik regresyon analizi mi, diskriminant analiz mi?

Log-lineer analiz

Çok değişkenli varyans analizi

Diğer çok değişkenli analiz yöntemleri

Faktör analizi

Küme analizi

SPSS'te çok değişkenli analizler nasıl yapılır?

Kovaryans analizi

Lojistik regresyon analizi

Diskriminant analiz

Log-lineer analiz

Çok değişkenli varyans analizi (MANOVA)

Daha önceki bölümlerin çoğunda iki değişkenli analiz yöntemlerini inceledik. Kolesterol düzeyi ile koroner arter çapı arasında bağıntı ve regresyon analizini yaparken, analiz konusu iki sayısal değişkendi. Miyokard infarktüsü sonrasında beta-bloker kullanımının kalsiyum antagonistleri ile karşılaşıldığı çalışmada iki nominal değişken arasındaki ilişki analiz edilmişti. Lösemide IL-3 düzeylerinin araştırıldığı çalışmada ise bir nominal ve bir sayısal değişken arasındaki ilişki analiz edilmişti.

Bu çalışmaların hepsinde değişkenlerden biri bağımsız ya da gruplandıracı değişken (kolesterol düzeyi, kullanılan ilaçın tipi ya da hastadan IL-3 düzeyi için kan alındığı dönemin tedavi öncesi ya da sonrası olması), diğer ise (koroner arter çapı, hastanın miyokard infarktüsü sonrası прогнозу ya da IL-3 düzeyi) bağımlı değişkendi.

İki değişkenli analiz yöntemlerinin uygulanması ve sonuçların yorumlanması oldukça kolay olmakla birlikte, her zaman gerçek yaşamla uyumlu sonuçları elde etmek mümkün olmayabilir. Çünkü biyolojik bilimlerde denek-

lerde ilgilenilen iki değişken dışındaki değişkenlerin değerlerinin sabit tutmak olanaksızdır. Bunu kolesterol düzeyi ve koroner arter çapı arasındaki ilişkinin araştırıldığı çalışmanın analizinden hatırlıyorsunuz.

Daha önceki bölümlerde birden çok bağımsız değişkenin olduğu analiz yöntemlerinden bazılarını incelemiştik. Örneğin iki-yönlü varyans analizi ya da çoklu regresyon analizinde bağımsız değişken sayısının birden çok olabileceğini görmüştük.

Bunlar dışında da çok değişkenli analiz teknikleri vardır. Hepsinde de hem hesaplamalar çok karışıktr (neyse ki artık bilgisayarlar var), hem de sonuçların yorumlanması iki değişkenli analizlerde olduğu gibi kolay değildir. Bu nedenle hemen hemen tüm çok değişkenli istatistik analiz teknikleri "ileri istatistik yöntemler" olarak ele alınmaktadır. Bu nedenle bu bölümde çok değişkenli analiz teknikleri hakkında fazla derine inmeden bilgi verilmekle yetinilecektir.

ÇOK DEĞİŞKENLİ ANALİZ YÖNTEMLERİNE GİRİŞ

Çok değişkenli yöntemlerin prototipi olarak çoklu regresyon analizinden söz etmekle işe başlamak gereklidir. Bölüm 14'te ayrıntılı olarak anlatılan çoklu regresyon analizinde, bildığınız gibi, bir sayısal değişkenin değeri, birden çok sayısal değişkenin değerlerine dayanılarak kestirilmeye çalışılmıştı. Hatırbağımsız değişkenlerden bir kısmının nominal değişken olması durumunda da çoklu regresyon analizi uygulanabileceğinden söz edilmiştir. Çoklu regresyon analizinde bir nominal değişkenin bağımsız değişken olarak kullanılabilmesi için önce "0" ve "1" olarak kodlanması gerektiğini de söylemiştim. Yine çoklu regresyon analizinde her değişkene karşılık gelen bir regresyon katsayısi ve bu katsayının 0'dan farklı olma olasılığını veren bir p değerinin hesaplandığını görmüştük. Çok değişkenli yöntemlerin çoğu da çoklu regresyon analizindeki bu kavramlar (nominal değişkenlerin "0" ve "1" olarak kodlanması gerekliliği, p değerinin anlamı gibi) ve terimler (regresyon katsayısi, regresyon eşitliği gibi) geçerlidir.

Çok değişkenli analiz yöntemlerinin ortak yanı, bir (ya da birden çok) bağımlı değişken üzerine etki etme olasılığı olan çok sayıda bağımsız değişkenin kendi aralarındaki etkileşimlerin ortaya çıkarılması ve böylece bağımsız değişkenlerin bağımlı değişken üzerine diğer değişkenlerin etkilerinden arıtlı, saf etkilerinin belirlenmesidir. Bu nedenle çok değişkenli analiz tekniklerinde birden çok bağımsız değişken analize alınabilir; ancak analizin uygulanabilmesi için denek sayısının da aynı oranda artırılması gereklidir. Denek sayısının kaç olması gerektiği hakkında belirli bir kural yoktur. Genel

olarak kabul gören görüş, denek sayısının bağımsız değişken sayısının yaklaşık on katı olması gerektidir. Örneğin koroner arter çapına kolesterol düzeyi, kan glukoz düzeyi, sistolik kan basıncı ve günde içilen sigara sayısının etkilerini incelemek istiyorsak, dört tane bağımsız değişken olduğuna göre 40 civarında hasta çalışılmalıdır.

Kovaryans analizi

ÖRNEK 75

Diabetes mellitusta (DM) eritrosit sedimentasyon hızının (ESH) normalden farklı olup olmadığını araştırmak amacıyla yapılan bir çalışmaya 30 tip I DM, 30 tip II DM ve 30 normal sağlıklı denek alınıyor ve ESH ölçülüyor. Yaş arttıkça ESH'nın de arttığı bilindiği için üç grubun yaş ortalamaları ve ESH ortalamaları tek yönlü varyans analizi ile karşılaştırılıyor ve Tablo 1'deki sonuçlar elde ediliyor.

Tablo 1'den ESH değerinin tip II DM'de yüksek, tip I DM'de düşük olduğu görülmektedir. Ama tip II DM grubunun yaş ortalaması da diğer grplara göre daha yüksek ve tip I DM grubunun yaş ortalaması ise daha düşüktür.

Tablo 1 Örnek 75'teki çalışma gruplarının yaş ve ESH ortalama \pm standart sapmaları ve tek yönlü varyans analizinden elde edilen p değerleri.

	Kontrol	Tip I DM	Tip II DM	p değeri
Yaş (yıl)	45 \pm 9	39 \pm 9	56 \pm 9	0.0001
ESH (mm/st)	34 \pm 7	30 \pm 8	41 \pm 7	0.0001

Bu durumda tip II DM grubunda ESH'nın yüksek olmasının nedeni tip II DM'e özgü nedenler midir, yoksa bu sonuç bu grubun daha yaşılı olmasına bağlı bir artefakt mıdır?

Bu soruyu yanıtlamadan önce çalışma gruplarında gerçekten yaş ile ESH arasında ilişki olup olmadığı incelenmelidir. İkişi de sayısal değişken olduğuna göre yaşın bağımsız, ESH'nm bağımlı değişken olduğu regresyon analizi yapıldığında aşağıdaki regresyon eşitlikleri ve regresyon katsayılarının önemliliğine karşılık gelen p değerleri elde edilmiştir (Şekil 1-3).

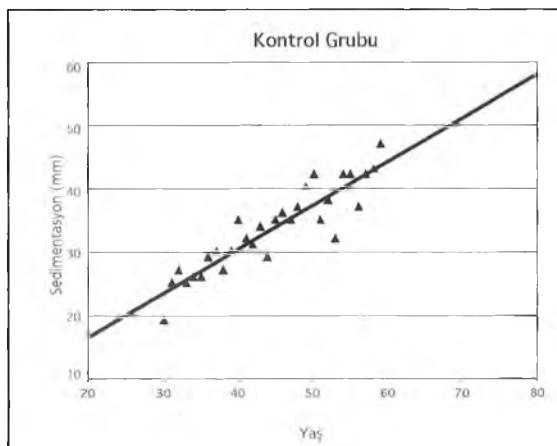
$$\text{Kontrol grubunda} \quad \text{ESH} = 2.83 + 0.70 \text{ (Yaş)} \quad p = 0.0001$$

$$\text{Tip I DM grubunda} \quad \text{ESH} = -0.46 + 0.78 \text{ (Yaş)} \quad p = 0.0001$$

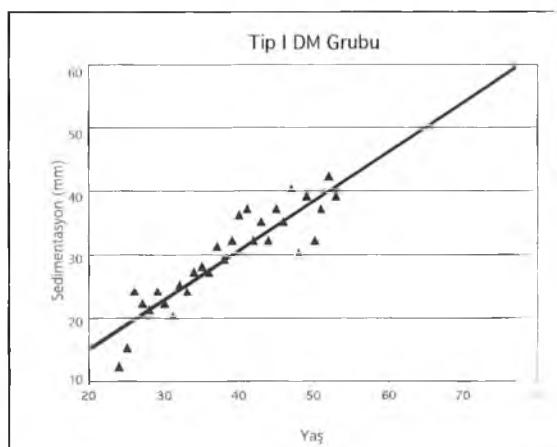
$$\text{Tip II DM grubunda } ESH = 1.52 + 0.71 \text{ (Yaş)} \quad p = 0.0001$$

$$\text{Tüm grup için } ESH = 2.18 + 0.70 \text{ (Yaş)} \quad p = 0.0001$$

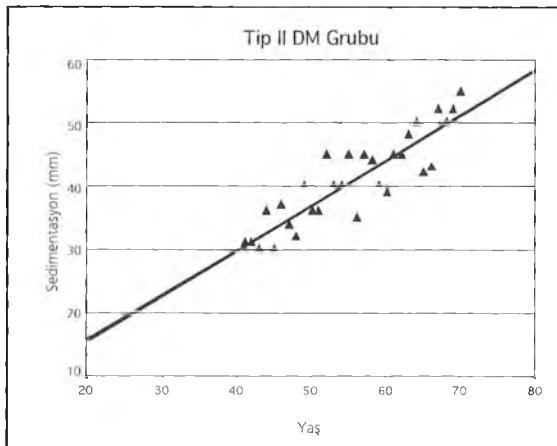
Göründüğü gibi tüm grplarda yaş arttıkça ESH artmaktadır. Araştırmanın amacı, yaşla ESH arasındaki ilişkiyi ortaya koymak değil, DM ile ESH arasındaki ilişkiyi ortaya koymaktır. Ama yine de yaş ile ESH arasındaki ilişkinin gözardı edilemeyeceği açıklar. Bu nedenle DM ile ESH ilişkisinden söz ederken, Tablo 1'de sonuçları verilen tek-yönlü varyans analizi yeterli değildir, çünkü yaşın ESH üzerine etkisi dikkate alınmadan yapılmıştır. Bu durumda uygulanması gereken analiz yöntemi kovaryans analizidir (*ANalysis of COVariance: ANCOVA*).



Şekil 1
Örnek 75'teki çalışmada kontrol grubunda yaş ile ESH arasındaki regresyon doğrusu.



Şekil 2
Örnek 75'teki çalışmada tip I DM grubunda yaş ile ESH arasındaki regresyon doğrusu.



Şekil 3

Örnek 75'teki çalışmada tip II DM grubunda yaş ile ESH arasındaki regresyon doğrusu.

Kovaryans analizi, iki ya da daha çok grupta bir sayısal değişkenin (ESH) ortalamalarının karşılaştırılması sırasında, bu değişkene etki eden başka bir sayısal değişkenin (yaş) etkisinin ortadan kaldırılması, başka bir deyişle grup etkisinin arıtılması amacıyla kullanılan çok yararlı bir istatistik analiz tekniğidir. Kovaryans analizinde, varyans analizindeki gibi “bağımsız” ve “bağımlı” değişkenler dışında, “araya giren” (*confounding*) bir ya da daha fazla değişken de vardır. Bu değişkene kovaryat (*covariate*) adı da verilir. Kovaryans analizinde tek bağımlı değişken olmasına karşın, birden fazla bağımsız değişken ve birden fazla kovaryat olabilir.

Kovaryans analizi, regresyon analizi ile varyans analizinin bir karışımıdır. Yukarıdaki örnekte bağımsız değişken olarak hastanın grubu, bağımlı değişken olarak ESH ve kovaryat olarak yaş alınarak, kovaryans analizi yapıldığında, ESH'nın yaşla arttığı ($p=0.0001$), yaşın etkisi arıtıldığında üç grup ESH ortalaması arasında fark olmadığı ($p=0.64$) anlaşılmıştır. Yine kovaryans analizi ile grupların ESH ortalamaları kovaryata göre düzeltilebilir. Buna göre düzeltilmiş ESH ortalamaları kontrol grubu için 35 mm/st , tip I DM grubu için 35 mm/st ve tip II DM grubu için 34 mm/st olarak hesaplanmıştır.

Kovaryans analizinin uygulanabilmesi için her grupta kovaryat ile bağımlı değişken arasındaki regresyon doğrularının eğiminin aynı olması koşulu vardır. Aksi takdirde, kovaryans analizi uygulanmamalıdır.

NOMİNAL DEĞİŞKENLERİN KESTİRİMİ

Çoklu regresyon analizi ve kovaryans analizinde bağımlı değişken olarak sayısal değişkenler vardır. Bağımlı değişkenin nominal olduğu durumlarda uygulanacak çok değişkenli analiz teknikleri de vardır. Bağımsız değişkenlerin tipi (nominal ya da sayısal) ve bağımlı nominal değişkenin alabildiği değer sayısına göre uygulanacak yöntem değişir.

Lojistik regresyon

ÖRNEK 76

Polikliniğe anemi bulguları ile başvuran her hastada rutin olarak hemoglobin (Hb), ortalama eritrosit hacmi (OEH), serum demir düzeyi, retikülosit ve periferik yayma yapılarak anemi nedeninin demir eksikliği olup olmadığı araştırılıyor.

Erkeklerde anemi saptandığında mutlaka ciddi bir yaklaşımla tanıya gitilmesi gereklidir. Kadınlarda ise aneminin altında yatan neden çoğunlukla kötü obstetrik ve jinekolojik geçmişe bağlı demir eksikliği anemisidir. Bu nedenle kadınlarda da erkeklerde olduğu gibi ayrıntılı inceleme hem para, hem de zaman ve emek açısından kayıplara yol açabilir. Anemi bulguları ile başvuran kadınların yaşı, çocuk sayısı, menopoz durumu, son gebeligidenden bu yana geçen süre, adet kanamalarının miktarı, beslenme öyküsü vb. bilgilerin de yardımcı ile serum demir düzeyi gibi pahalı ve retikülosit ve periferik yayma gibi zaman ve emek alıcı incelemeler yapılmadan da doğru tanı konup konulamayacağının saptanması amacıyla yapılan bir çalışmada, 50'si demir eksikliği, 50'si diğer nedenlere bağlı anemisi olan 100 kadın hastada ayrıntılı öykü ve fizik inceleme bulguları kaydediliyor ve yukarıdaki laboratuvar incelemeler yapılıyor. Öykü ve fizik incelemektedeki bulgular ve Hb ve OEH'ne dayanarak, hastaların "demir eksikliği anemisi" ya da "diğer anemiler" gruplarından hangisine girdiğinin doğru olarak saptanıp saptanamayacağı inceleniyor.

Bu örnekteki çalışma, lojistik regresyon için prototip olarak alınabilecek bir sorundur. Bu çalışmanın analizine başlamadan önce değişkenleri tanımlayalım. İki basitleştirmek için değişkenleri az sayıda tutalım. Her hastanın OEH ve yaş olmak üzere iki sayısal değişken ve menopoz durumu (premenopozal=1, postmenopozal=0) ve adet durumu (fazla kanama=1, normal miktarda kanama ya da postmenopozal=0) olmak üzere iki nominal değişkenin değerlerinin bilindiğini düşünelim. Doğal olarak bir diğer değişken de yapılan tüm incelemeler sonucunda hastanın ait olduğu gruptur; yani "demir eksikliği anemisi" ve "diğer anemiler" olarak iki değer alabilen bir nominal değişkendir. Bu değişkene "anemi nedeni" adını verelim. "Demir eksik-

liği anemisi=1" ve "diğer anemiler=0" olarak kodlanmış olsun.

Basit istatistik analiz yöntemleri ile Örnek 76'daki sorunun çözümünü arayalım. Örneğin iki grubun OEH ve yaş ortalamaları arasında fark olup olmadığı Student t testi ile incelendiğinde, "demir eksikliği anemisi" grubunun diğer gruba göre daha genç (37 yaş ve 51 yaş) ve OEH ortalamasının daha düşük (66 fL ve 89 fL) olduğu saptanıyor. Buna dayanarak, mikrositer anemisi olan genç bir kadında aneminin nedeninin demir eksikliği olma olasılığının fazla olduğunu belki sezebiliriz, ama Student t testi ile elde edebildiğimiz tek yorum demir eksikliği anemisi grubunun OEH ve yaş ortalamasının diğer gruptan daha düşük olduğundan daha fazla bilgi içermez. Üstelik menopoz ve adet durumu ile anemi nedeni arasındaki ilişki tamamen gözardı edilmişdir.

Anemi nedenine ek olarak menopoz ve adet durumunun da bağımsız faktörler olarak alındığı iki-yönlü varyans analizi yapılrsa, menopoz ve adet durumu ile anemi nedeni arasında etkileşim bulunacağı çok olasıdır ve bu durumda iki-yönlü varyans analizi geçerli olmayacağıdır. Etkileşim olmasa bile bulunacak sonuç anemi gruplarının OEH ve yaş ortalamaları arasında fark olup olmadığıdır ya da yalnızca bir grubun ortalamasının diğer gruptan farklı olduğunu söyleyebiliriz. Serum demir düzeyi, retikülosit ve periferik yayma incelemesi yapılmadan doğru tanı konulup konulamayacağına ilişkin bilgi elde edilmemiştir.

Öte yandan menopoz durumu ve adet durumu ile aneminin nedeni arasında ilişki olup olmadığını uygun bir çapraz tablo istatistiği ile inceleyebiliriz. Ki-kare analizi ile demir eksikliği anemisi olan kadınların %74'ünün pre-menopozal dönemde ve %62'sinin adet kanamasının fazla olması ve diğer grupta ise bu oranların sırasıyla %16 ve %8 olarak saptanması ve iki grup oranı arasındaki farkın istatistiksel olarak önemli olması da bir fikir verir, ama yine yalnız bu verilere dayanarak, doğru tanı konup konamayacağını bileyemeyiz. Ayrıca, bu kez de OEH ve yaş ile anemi nedeni arasındaki ilişki dikkate alınmamıştır.

Bağımlı yani diğer değişkenlerin değerleri bilindiğinde kestirilecek değişken burada olduğu gibi nominal bir değişken (anemi nedeni) değil de, sayısal bir değişken (örneğin serum demir düzeyi) olsaydı, yani dört bağımsız değişkenin değerleri bilindiğinde serum demir düzeyinin kestirilip kestirilemeyeceği çalışmanın konusu olsaydı, çoklu regresyon analizi ile sorun çözülecekti. Oysa çoklu regresyon analizinde animsanacağı gibi, bağımlı değişkenin sayısal değişken olması vazgeçilmez bir koşuludur.

Analitik istatistik tekniklerin başlıca iki amaçla kullanıldığından, bunlardan birinin fark olup olmadığını incelenmesi, diğerinin ise ilişki saptama ve kestirim yapabilme gücünün test edilmesi olduğundan söz edilmiştir. Örnek

76'daki çalışmada, iki grup ortalaması ya da oranı arasında fark olup olmadığını, Student t testi, ki-kare analizi ya da regresyon analizi ile saptanmasının bizi istediğimiz sonuca götürmediği görülüyor. Öte yandan kestirim yapabilmek için kullanılan çoklu regresyon analizini ise uygulamak olası değil. O halde çözüm nedir?

Bu sorun lojistik regresyon analizi ile çözülebilir. Lojistik regresyonda da, regresyon analizinde olduğu gibi bazı değişken değerlerine dayanarak kestirim yapılmaya çalışılır, ancak iki yöntem arasında üç önemli fark vardır:

1. Regresyon analizinde kestirilecek olan (yani bağımlı) değişken sayısal iken, lojistik regresyonda bağımlı değişken ikili (dikotom) nominal bir değişken olmalıdır.
2. Regresyon analizinde bağımlı değişkenin değeri, lojistik regresyonda ise bağımlı değişkenin alabileceği iki değerden birinin gerçekleşme olasılığı kestirilir. Örneğin 43 yaşında, adet kanaması fazla olan ve OEH 67 fL olan bir kadının serum demir düzeyinin 40 mg/dL olduğu çoklu regresyon analizi ile, bu hastada anemi nedeninin demir eksikliği olması olasılığının %87 olduğu lojistik regresyon analizi ile kestirilebilir.
3. Regresyon analizinde bağımsız değişkenlerin çoklu normal dağılım gösternesi koşulu aranırken, lojistik regresyon uygulanabilmesi için bağımsız değişkenlerin dağılımına ilişkin hiç bir ön koşul gerekmeyez.

Örnek 76'daki çalışmada SPSS ile lojistik regresyon analizi yapıldığında, Şekil 4'teki gibi bir tablo elde edilecektir. Bu tabloda B ile gösterilen rakamlar de her değişkene ait regresyon katsayılarıdır. Sig sütunundaki değerler ise, hastaların iki gruptan hangisine girdiğinin belirlenmesinde değişkenlerin katkısının istatistiksel olarak önemli olup olmadığını gösteren p değerleridir. Şekil 4'den OEH'nin bilinmesi ile hastaların gruplandırmasının kolaylaşacağı ($p=0.0008$), menopoz durumu ($p=0.20$) ve adet durumunun ($p=0.31$) bilinmesinin katkısının önemli olmadığı anlaşılmaktadır. Yaşın istatistiksel önemi ise sınırdadır ($p=0.052$). Göründüğü gibi Student t testi ile iki grubun yaş ortalamaları ve ki-kare analizi ile premenopozal ya da fazla kanaması olan kadınların oranları arasındaki farkın istatistiksel olarak önemli bulunmasına karşın, hastaların "demir eksikliği anemisi" ya da "diğer anemiler" grubundan hangisine girdiğinin kestirilmesinde fazla katkısı olmadığı bulunmuştur. Çelişki gibi görünen bu durumun nedeni, bağımsız değişkenler arasındaki karşılıklı bağımlılıklardır.

Variables in the Equation							
Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
YAS	-.1573	.0811	3.7614	1	.0524	-.1127	.8545
OEH	-.4120	.1231	11.2077	1	.0008	-.2577	.6623
MENOPOZ(1)	1.8548	1.44141	.6558	1	.1982	.0000	6.3905
ADET(1)	-1.9023	1.8927	1.0101	1	.3149	.0000	.1492
Constant	38.6833	12.0726	10.2670	1	.0014		

Şekil 4 Lojistik regresyon modeline alınan değişkenler, regresyon katsayıları ve önemlilik derecelerinin yer aldığı tablo.

Şekil 4'de R ile gösterilen rakamlar bağımlı değişkenle bağımsız değişken arasındaki bağıntı katsayısıdır. Bağıntı analizinde olduğu gibi R değeri -1 ile +1 arasında olabilir. P değeri küçüldükçe, R değeri -1'e ya da +1'e yaklaşır. Pozitif R değeri, bağımsız değişkenin değeri arttıkça, bağımlı değişkenin olasılığının artlığına işaret eder. R değeri negatifse, bağımlı değişkenin olasılığı azalmaktadır. Örneğin OEH'nin R değeri negatif işaretli olduğu için, OEH arttıkça, demir eksikliği olasılığının azaldığı anlaşılmaktadır. (Lojistik regresyon analizinde bağımlı değişkene 0 ve 1 kodları verilir ve kodu 1 olan durumun olasılığı hesaplanır. Bu nedenle "demir eksikliği anemisi=1" ve "diğer anemiler=0" olarak kodlanmıştır.)

Lojistik regresyon analizi ile hesaplanan olasılık %50'den büyükse, aneminin nedeninin demir eksikliği, %50'den küçükse diğer nedenler olduğu varsayımlı ile hastaların ne kadar doğrulukla sınıflandırılabileceği de SPSS'te tablo halinde elde edilebilir (Şekil 5). Sınıflandırma tablosu (*classification table*) adı verilen bu tabloda her iki grup için ayrı ayrı ve tüm grup için doğru ve yanlış sınıflandırılan hasta sayıları ve doğru sınıflandırma oranları verilmiştir. "1" ile kodlanmış olan demir eksikliği grubundaki 50 hastadan 49'unun (%98), "0" ile kodlanmış diğer gruptaki 50 hastadan 48'inin (%96) doğru olarak sınıflandırıldığı anlaşılmaktadır. Genel doğruluk oranı ise %97'dir.

Classification Table for ANEMI			Predicted			Percent Correct
			Diğer anemiler	Demir eksikliği		
Observed		0	I	1		96.00%
		+-----+ Diğer anemiler 0 I 48 I 2 I +-----+ +-----+				
Demir eksikliği	1	I 1 I 49 I +-----+ +-----+				98.00%
				Overall		97.00%

Şekil 5 Lojistik regresyon modelinin ne kadar başarılı olduğu hakkında fikir veren gösterim şekillerinden birisi olan “Sınıflandırma Tablosu”.

Sonuç olarak, lojistik regresyon analizi, bir durumun gerçekleşme ya da gerçekleşmemeye olasılığını, sayısal ya da nominal olsun birden çok değişkeni ölçerek hesaplamak ya da kestirmek amacıyla kullanılan bir istatistik analiz yöntemidir.

Tıpta başlıca kullanım amaçları arasında hastalık risk etkenlerinin belirlenmesi ve tek tek bireylerin belirli özelliklerini bilindiğinde ya da bireylerde belirli ölçümler yapıldığında hastalık olasılığının kestirilmesi için modeller geliştirilmesi yer alır.

Diskriminant analiz

Diskriminant analizde de lojistik regresyonda olduğu gibi hastaların hangi gruba ait olduğunu kestirmek için hangi değişkenlerin önemli kestiriçi (*predictor*) olduğu araştırılır; ya da başka bir tanımlamayla, birden çok değişkenin değerleri bilindiğinde hastaların bir gruba ait olma olasılıkları hesaplanabilir.

Diskriminant analizin lojistik regresyondan iki önemli farkı vardır:

1. Lojistik regresyon uygulanabilmesi için bağımsız (*predictor*) değişkenlerin dağılım özelliklerine ilişkin hiç bir koşul olmamasına karşın, diskriminant analizde bağımsız değişkenlerin çoklu normal dağılıma uyması koşulu aranır. Bu nedenle sayısal bağımsız değişkenlerin dağılım özelliklerini dikkate alınmalıdır, nominal değişkenlerin ise bağımsız değişken olmak kullanılması hatalı değildir, ancak sakıncalı olabilir.
2. Diskriminant analisin lojistik regresyondan farklı olan bir diğer özelliği ise

hastaların ait olabilecekleri grup sayısının iki ile kısıtlı olmamasıdır. Yani lojistik regresyonda bağımlı değişkenin ikili nominal (dikotom) değişken olmasının kesin koşul olmasına karşın, diskriminant analizde ikiden çok değer alabilen nominal ya da ordinal bir değişken bağımlı değişken olabilir.

ÖRNEK 77

Örnek 76'daki çalışmada hastaların "demir eksikliği anemisi" ve "diğer anemiler" gruplarından hangisine girdiğini belirlemek için öykü ve fizik muayenedeki bulgular ve Hb ve OEH ölçümlerinin yeterli olup olmayacağı araştırılmıştı. Başka bir çalışmada hastaların "demir eksikliği anemisi", "talassemi" ve "diğer anemiler" gruplarından hangisine girdiğinin kestirilmesi amacıyla Örnek 76'daki çalışma düzeni kullanılıyor ve bu kez her grupta 50 hasta olmak üzere 150 hastada inceleme yapılmıyor. Her bir hastanın bulgularına ve laboratuvar değerlerine dayanarak üç gruptan hangisine girdiğinin doğru olarak kestirilip kestirilemeyeceği araştırılıyor.

Örnek 76'daki çalışmada, hastalar iki gruba ayrıldığı için lojistik regresyon analizi ile sorun incelenmişti. Örnek 77'deki çalışmada ise hastaların "demir eksikliği anemisi", "talassemi" ve "diğer anemiler" gruplarından hangisine girdiği kestirilmek isteniyor. Grup sayısı ikiden fazla olduğu için, yani bağımlı değişken artık ikili nominal değişken olmadığı için lojistik regresyon analizi uygulanamaz. Bu durumda uygulanacak yöntem diskriminant analizidir.

Diskriminant analizde de hangi bağımsız değişken ya da değişkenlerin, hastaların gruplandırılabilmesi için yararlı olduğu ve her bir hastanın hangi gruba ait olduğunu belirlenebilmesinin bu değişkenlere dayanarak ne kadar başarılı olduğu hesaplanabilir. Yine diskriminant analizle de sınıflandırma tablosu elde edilebilir. Şekil 6'da Örnek 77'deki çalışmada verilere diskriminant analiz uygulandığında hastaların ne kadarının doğru olarak gruppalandırıldığı görülmüür. Birinci satırındaki rakamlar "diğer anemiler" grubundaki 50 hastanın diskriminant analizi ile nasıl sınıflandırıldığı yer alıyor. Bu hastaların 46'sında (%92) doğru tanı konulabilenken, 4'ünün "demir eksikliği anemisi" olduğu kestirilmiş. Diğer iki grupta da hastaların nasıl gruppalandırıldığı tabloda görülmüür. Genel doğruluk oranı ise %76.7.

		Classification Results ^a			
		Predicted Group Membership			
		Düger anemiler	Demir eksikliği anemisi	Talassemi	Total
Original Count	Anemi nedeni	46	4	0	50
	Düger anemiler	6	35	9	50
	Demir eksikliği anemisi	0	16	34	50
%	Düger anemiler	92.0	8.0	.0	100.0
	Demir eksikliği anemisi	12.0	70.0	18.0	100.0
	Talassemi	.0	32.0	68.0	100.0

a. 76.7% of original grouped cases correctly classified.

Şekil 6 Diskriminant analiz modelinin ne kadar başarılı olduğu hakkında fikir veren gösterim şekillerinden birisi olan "Sınıflandırma Tablosu".

Lojistik regresyon analizi mi, diskriminant analiz mi?

Lojistik regresyon analizi ile diskriminant analiz arasındaki yakınlık nedeniyle, bazen bu iki yöntemden hangisinin kullanılacağı konusunda karar vermekte güçlük çekilebilir. Bu durumda doğru seçim yapabilmek için şunu noktaları dikkate almak yararlı olacaktır:

1. Bağımlı nominal değişken ikili (dikotom) değilse, yani deneklerin ikiden çok gruptan hangisine girdiği kestirilecekse, lojistik regresyon analizi uygulanamaz. Bu durumda diskriminant analiz seçilmelidir. Bağımlı değişken ikili nominal değişkense, her iki yöntem de uygulanabilir.
2. Lojistik regresyon analizi, bağımsız değişkenlerin dağılımı ne olursa olsun uygulanabilir. Oysa diskriminant analizde bağımsız değişkenlerin çoklu normal dağılım göstermesi koşulu aranır. Bu nedenle bağımsız değişkenlerin çoğu nominalse, lojistik regresyon analizi seçilmelidir. Bağımsız değişkenlerin çoğu sayısalsa, her iki yöntem de kullanılabilir, ancak yine diskriminant analiz için bu değişkenlerin dağılım özellikleri incelenmelidir.

Son söz olarak, tipten hem lojistik regresyon analizi, hem de diskriminant analizin giderek artan sıklıkla kullanıldığını söylemek gereklidir. Çünkü her iki yöntemde de, yalnızca çalışma grubuna ilişkin bir sonuçla yetinilmesi gerekmek. Bir çalışmanın analizinden elde edilen bir denklemden, daha sonraki bireylerde de hastalık, ölüm vb. olayların olasılığını kestirebilmek mümkün olur.

Örnek 76'daki çalışmada yalnızca iki grup OEH ortalaması arasındaki farkın istatistiksel olarak önemli olduğu sonucuna ulaşılan Student t testi ile yetinseydik, daha sonra izleyeceğimiz hastalarda bu çalışmanın sonuçlarını

dayanarak, hastada ne tür bir anemi olduğunu belirlemekte elimizde objektif bir kriter olmayacağı. Oysa, lojistik regresyon analizi ile elde ettiğimiz denklemi kullanarak, daha sonraki anemik hastalarda da anemi türünü kesitmek mümkün olacaktır.

Log-lineer analiz

Bağımsız grup oranlarının karşılaştırılmasında kullanılan çapraz tablo istatistikleri Bölüm 12'de ayrıntılı olarak incelenmişti. Bu yöntemleri, başka bir tanımlamayla belirtirsek, iki nominal değişken arasında ilişki olup olmadığıının incelenmesine yarayan yöntemler olarak tanımlayabiliriz. Bölüm 12'de verilen Örnek 76'daki çalışmanın analizinde ki-kare yöntemi ile elde edilen sonuç iki şekilde dile getirilebilir: "İki tedavi grubundaki ölüm oranları arasında fark vardır" ya da "İlaç değişkeni ile прогноз değişkeni arasında ilişki vardır". İşte çapraz tablo istatistiklerinde ilgi konusu iki nominal değişken arasında ilişki olup olmadığıdır.

İkiden çok nominal değişken olduğu zaman akla ilk gelen birkaç tane çapraz tablo hazırlayıp, tek tek her tablonun istatistiklerini yapmaktadır.

ÖRNEK 78

Beta-bloker ve kalsiyum antagonistlerinin anstabil anjina pektoriste miyokard infarktüs (MI) gelişmesini önlemek üzerindeki etkilerinin karşılaştırılması

amacıyla düzenlenen bir çalışmaya alınan 80 hasta iki gruba ayrılıyor ve bir gruba beta-bloker, diğer gruba kalsiyum antagonisti veriliyor. Daha sonraki 6 ay içinde hastaların ne kadarının MI geçirdiği gözleniyor.

Örnek 78'deki çalışmada uygulanması gereken analiz yönteminin ki-kare analizi olduğu görülmüyor. Tablo 2'den Tablo 7'ye kadar yer alan tablolarda bu çalışmada verilere ilişkin çapraz tablolar görülmektedir. Beta-bloker grubunda 21, kalsiyum antagonisti grubunda 30 hastanın MI geçirdiği gözleniyor ve ki-kare analizi ile beta-bloker grubundaki MI oranının kalsiyum antagonisti grubundan istatistiksel olarak önemli derecede az olduğu görülmüyor ($p=0.03$) (Tablo 2).

Tablo 2 Örnek 78'deki çalışmada iki tedavi grubunda MI geçiren hasta sayılarının gösterildiği çapraz tablo.

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

Ancak gruptardaki hastaların cinsiyetlere göre dağılımının farklı olduğu görülüyor. Beta-bloker grubunda 10 erkek, 30 kadın, kalsiyum antagonist grubunda 30 erkek, 10 kadın var (Tablo 3). 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 прогноз arasındaki ilişki açısından incelendiği zaman her iki grupta da erkeklerde MI oranının daha yüksek olduğu saptanıyor (Tablo 4 ve 5).

Tablo 3 Örnek 78'deki çalışmada iki tedavi grubundaki hastaların cinsiyete göre dağılımının gösterildiği çapraz tablo.

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

Tablo 4 Örnek 78'deki çalışmada beta-bloker grubunda farklı cinslerde MI geçiren hasta sayılarının gösterildiği çapraz tablo.

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

Tablo 5 Örnek 78'deki çalışmada kalsiyum antagonist grubunda farklı cinslerde MI geçiren hasta sayılarının gösterildiği çapraz tablo.

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

Acaba beta-bloker grubunda infarktüs oranının düşük olmasının nedeni ilaçın 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 bir çözümü, tedavi gruplarındaki infarktüs oranlarının her iki cins için ayrı olarak incelenmesi olabilir. Gerçekten de hem erkeklerde (Tablo 6), hem de kadınlarda (Tablo 7) hangi ilaç kullanılırsa kullanılsın, прогнозun farklı olmadığı görülmektedir.

Tablo 6 Örnek 78'deki çalışmada iki tedavi grubunda MI geçiren erkek hasta sayılarının gösterildiği çapraz tablo.

	Miyokard İnfarktüsü		Toplam
	+	-	
Beta-bloker	8	2	10
Kalsiyum antagonisti	27	3	30
Toplam	35	5	40

Tablo 7 Örnek 78'deki çalışmada iki tedavi grubunda MI geçiren kadın hasta sayılarının gösterildiği çapraz tablo.

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

Örnek 78'deki çalışmanın analizini 6 tane çapraz tablo ile yapabildik. Yapılan işlemi tanımlarsak, tedavide kullanılan ilaç, cinsiyet ve прогноз olmak üzere üç nominal değişken arasındaki ilişkiyi 6 adet ki-kare testi ya da Fisher testi ile çözümledik. Değişken sayısı üç değil de, beş, on, hatta daha fazla olabilir. Bu durumda belki yüzlerce çapraz tablo yapıp, tek tek hepsine istatistik analiz uygulamak gerekecekti.

İkiden çok nominal değişken arasındaki ilişkiyi incelemenin doğru yolu değişkenler arasındaki ilişkiyi ikişer ikişer olarak ki-kare analizleri ile incele-

mek değil, log-lineer analiz yapmaktadır.

Log-lineer analiz yönteminin yararı yalnızca fazla sayıda çapraz tablo yapmanın zahmetinden kurtulmak değil, Örnek 78'deki cinsiyet-prognоз ilişkisi gibi gözden kaçabilecek etkileşimleri ortaya koyup, değişkenler arasındaki gerçek ilişkileri saptamaktır.

Ki-kare analizinde olduğu gibi, log-lineer analizde de değişkenler bağımlı ve bağımsız olarak sınıflandırılmazlar.

Örnek 78'deki çalışma, log-lineer analizle incelendiğinde, iki tedavi grubundaki hastaların cinsiyet dağılıminin benzer olmadığı ($p<0.0001$), prognозun iki cinste benzer olmadığı ($p=0.0001$), ancak tedavi ile prognоз arasında ilişki olmadığı ($p=0.88$) saptanmıştır.

ÇOK DEĞİŞKENLİ VARYANS ANALİZİ

Çok değişkenli varyans analizi (*Multivariate ANalysis Of VAriance: MANOVA*) bağımlı değişken sayısının birden fazla olduğu bir varyans analizidir.

Örneğin tip I ve tip II diabetes mellituslu (DM) hastalar ve sağlıklı hastalarda trigliserid, HDL-kolesterol ve LDL-kolesterol düzeyi ortalamalarının farklı olup olmadığını ve cinsiyetin bu değişkenler üzerinde etkili olup olmadığını inceleyen, deneğin grubu (tip I DM, tip II DM ya da normal) ve cinsiyetini bağımsız, trigliserid düzeyini bağımlı değişken olarak alıp, iki-yönlü varyans analizi uygulayabilir ve bu işlemi HDL-kolesterol ve LDL-kolesterol için yineleyebiliriz.

Bunun yerine yapılması gereken, bu üç değişkeni bağımlı değişkenler, grup ve cinsiyet değişkenlerini de bağımsız değişkenler olarak aldığımız çok değişkenli varyans analizi uygulamaktır. Eğer bağımlı değişkenler üzerine etkisi olduğu düşünülen sayısal değişkenler de varsa, bunların kovaryat olarak modele alınması ile çok değişkenli kovaryans analizi (*Multivariate Analysis of COVariance: MANCOVA*) yapılabilir.

Tek tek varyans analizi yapmak yerine çok değişkenli varyans analizi yapmanın asıl nedeni, tek tek t testi yapmak yerine bir tek varyans analizi yapmactaki nedenin aynısıdır. İkiden çok grup ortalamasının karşılaştırıldığında varyans analizi yerine, grupların ikişer ikişer karşılaşması ile gerçekte istatistiksel olarak önemli fark olmadığı halde, varmış gibi sonuçların elde edilebileceğini biliyoruz.

Birden çok bağımlı değişken olduğunda, çok değişkenli varyans analizi yerine, her değişken için ayrı bir tek yönlü varyans analizi yapmakta da ayırı sakınca vardır. Bu durumda yapılması gereken önce tüm bağımlı değişkenlerin toplu olarak çok değişkenli varyans analizine alınmasıdır. Örneğin

deneğin grubu ve cinsiyetini bağımsız değişken, trigliserid, HDL-kolesterol ve LDL-kolesterol değerlerini bağımlı değişken olarak aldığımızda, her bağımsız değişken için bir tek p değeri (*significance of multivariate test*) elde edilir. Örneğin cinsiyete karşılık gelen p değeri 0.16 ise, erkek ve kadınların ortalaması trigliserid, HDL-kolesterol ve LDL-kolesterol değerlerinin farklı olmadığı anlaşıılır. Bu durumda tek tek varyans analizi yapmaya gerek kalmaz. Eğer *multivariate* p değeri 0.05'ten küçükse, üç bağımlı değişkenden hangisi ya da hangilerinin erkek ve kadın hastalarda farklı olduğunu ortaya koymak için üç ayrı tek-yönlü varyans analizi yapılabilir. Ya da daha doğrusu çok değişkenli varyans analizi yönteminin bir parçası olarak her bağımlı değişken için hesaplanabilecek tek değişken için p değeri (*significance of univariate test*) kullanılabilir.

Örneğin cinsiyet değişkeninin *multivariate* p değeri 0.05'ten büyükse, *univariate* p değerlerinin de 0.05'ten büyük olması beklenir. Eğer *multivariate* p değeri 0.05'ten küçükse, erkek ve kadın hastalarda hangi bağımlı değişkenin ortalamasının farklı olduğunu anlamak için *univariate* p değerlerine dikkat etmek gereklidir. *Univariate* p değeri 0.05'ten küçük olan bağımlı değişkenin ortalamasının iki grupta farklı olduğu anlaşıılır.

Şekil 7'de SPSS ile elde edilmiş çok değişkenli regresyon analizinin *multivariate* ve *univariate* p değerleri görülmektedir.

Tip değişkeni için *multivariate* p değerinin 0.05'ten, ve gruplardaki HDL-kolesterol ortalamalarına karşılık gelen *univariate* p değerinin 0.05'ten küçük olduğu görülmektedir. Yani tip I DM, tip II DM ve normal gruplarda ortalaması HDL-kolesterol değeri farklı, trigliserid ve LDL-kolesterol değerleri ise farklı değildir. Cinsiyet değişkenine karşılık gelen *multivariate* p değeri ise 0.79'dur.

Trigliserid, HDL-kolesterol ve LDL-kolesterolle karşılık gelen p değerlerinin de 0.05'ten büyük olduğu görülmektedir. Yani ölçülen her üç değişkenin de ortalaması erkek ve kadın hastalarda farklı değildir.

Multivariate Tests				
Effect		Value	F	Sig.
Intercept	Pillai's Trace	.919	173.945	.000
	Wilks' Lambda	.081	173.945	.000
	Hotelling's Trace	11.344	173.945	.000
	Roy's Largest Root	11.344	173.945	.000
TIP	Pillai's Trace	.322	3.005	.010
	Wilks' Lambda	.684	3.205	.007
	Hotelling's Trace	.453	3.397	.005
	Roy's Largest Root	.433	6.777	.001
CINSIYET	Pillai's Trace	.023	.354	.787
	Wilks' Lambda	.977	.354	.787
	Hotelling's Trace	.023	.354	.787
	Roy's Largest Root	.023	.354	.787
TIP * CINSIYET	Pillai's Trace	.082	.670	.674
	Wilks' Lambda	.918	.666	.677
	Hotelling's Trace	.088	.661	.681
	Roy's Largest Root	.081	1.274	.294

Tests of Between-Subjects Effects				
Source	Dependent Variable	F	Sig.	
Corrected Model	Trigliserid	.825	.538	
	HDL-kolesterol	4.283	.003	
	LDL-kolesterol	.348	.881	
Intercept	Trigliserid	83.281	.000	
	HDL-kolesterol	99.379	.000	
	LDL-kolesterol	341.202	.000	
TIP	Trigliserid	1.009	.372	
	HDL-kolesterol	9.975	.000	
	LDL-kolesterol	.533	.590	
CINSIYET	Trigliserid	.390	.535	
	HDL-kolesterol	.026	.873	
	LDL-kolesterol	.017	.898	
TIP * CINSIYET	Trigliserid	1.085	.346	
	HDL-kolesterol	.225	.799	
	LDL-kolesterol	.257	.775	

Şekil 7
SPSS ile elde edilen çok değişkenli varyans analizi tablosu.

DİĞER ÇOK DEĞİŞKENLİ ANALİZ YÖNTEMLERİ

Bu bölümde şimdije dek sözü edilen istatistik yöntemlerinin ileri istatistik analiz yöntemleri olduğunu vurgulamıştık. Faktör analizi, küme (*cluster*) analizi, varyans komponent analizi (*variance components analysis*) gibi profesyonel yöntemler de sık olmasa da tip literatüründe karşımıza çıkmaktadır. Bu analizlere ilişkin ayrıntıya girmekten kaçınarak, bir iki cümle ile tanımla-

makla yetinilecektir. Çünkü bu yöntemlerin kesinlikle deneyimli bir istatistik uzmanının yardımı ile uygulanması gereklidir.

Faktör analizi

Faktör analizinde diğer analiz yöntemlerinde alışık olduğumuz gibi bağımlı ve bağımsız değişkenler yoktur. Değişkenlerin hepsi bağımsız değişkendir.

Faktör analizi, bir grup hastada ölçülen sayısal değişkenler arasındaki ilişkiye inceleyerek, hastalardaki farklılıklar (varyasyonu) en çok belirleyen değişkenlerin ortaya konulmasına yarar. Başka bir tanımlamayla, çok sayıda değişkenden hastalardaki farklılığı en iyi özetleyen değişkenlerin belirlenmesi faktör analizi ile yapılır.

Tıpta sık kullanıldığı alanlardan biri, pilot çalışmalarında anketlere verilen yanıtların değerlendirilmesi ile anket formlarının daha sade hale getirilmesi ve aynı amaca hizmet eden soruların tekrar tekrar sorulmasının önlenmesinin sağlanmasıdır. Ya da ankete verilen yanıtların güvenilirliğini test etmek amacıyla aynı anlama gelen soruların neler olduğunu belirlemesinde de faktör analizi yardımcı olur. Faktör analizindeki değişkenler sayısal olabileceği gibi, ordinal de olabilir. Hasta sayısının, değişken sayısının en az 10 katı kadar olması gereklidir. Örneğin 25 soru içeren bir anketin değerlendirilmesinde faktör analizinin kullanılabilmesi için anketin en az 250 deneğe uygulanması gereklidir.

Küme analizi

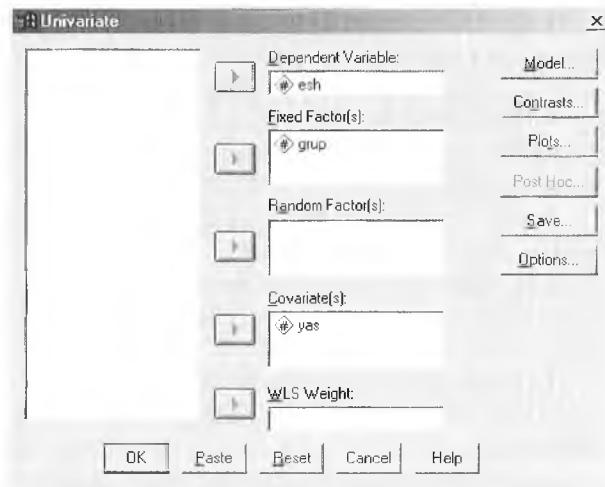
Küme analizinde de, faktör analizi gibi, bağımlı ve bağımsız değişkenler yoktur. Değişkenlerin hepsi bağımsız değişkendir. Küme analizi, bir açıdan faktör analizine, diğer açıdan diskriminant analize benzer. Faktör analizinde değişkenler arasındaki ilişkinin derecesi ya da birbirlerine yakınlığı incelenirken, küme analizinde hastaların birbirine yakınlığı incelenir. Hastalar özelliklerine göre gruplara, ya da kümelere ayrılır. Diskriminant analizde benzerliği ne gelince, küme analizinin mantığı, hastaların ait olduğu grupların baştan belli olmadığı bir diskriminant analizi gibidir.

SPSS'TE ÇOK DEĞİŞKENLİ ANALİZLER NASIL YAPILIR?

Kovaryans analizi

Analyze >> General Linear Model >> Univariate

Yukarıdaki seçimle **Univariate** diyalog kutusu açılır (Şekil 8). Burada kırı laştırılacak ortalamaları belirleyen değişken, yani bağımlı değişken *Dependent*



Şekil 8
SPSS'te "Univariate"
diyalog kutusu.

Variable kutusuna, grupları belirleyen değişkenler, yani faktör ya da bağımsız değişkenler *Fixed Factor(s)* kutusuna, kovaryat *Covariate(s)* kutusuna seçilmelidir. Şekil 8'de Örnek 75'teki çalışmanın analizi için uygun değişken seçiminin yapıldığı görülmektedir.

Analyze başlamadan önce *Contrasts* altmenüsüne girip *Change Contrast*'tan *Simple*'ı (basit kontrast) ve *Reference Category* olarak da *First*'ü seçin. *Change* butonuna basarak grup değişkeninin kontrast tipini değiştirebilir. Böylece ikişerlik karşılaştırımları 0 ile kodlanmış olan kontrol grubu ile 1 ve 2 ile kodlanmış olan tip I DM ve tip II DM grupları arasında yapılacaktır.

Tanımlayıcı istatistikleri almak için *Options* altmenüsüne girip *Descriptive statistics*'i seçin. Kovaryat etkisinden arıtlılmış ortalamaları almak için *Factor(s) and Factor Interactions* listesinden grup değişkenini *Display Means for* listesine seçin.

Şekil 9'da Örnek 75'teki çalışmada kontrol grubu, tip I DM ve tip II DM grupları arasında yaşın etkisinden arıtlılmış olarak sedimentasyon hızı ortalamalarının karşılaştırıldığı kovaryans analizinin sonuçları görülmektedir.

Tests of Between Subjects Effects tablosunda grup değişkenine karşılık gelen p değerinin 0.64 olduğu görülmektedir, yani üç grup arasında yaşın etkisinden arıtlılmış olarak sedimentasyon hızı ortalamaları farklı değildir. *Contrast Results* tablosunda kontrol grubu ile tip I DM arasındaki farkın p değerinin 0.73 ve kontrol grubu ile tip II DM arasındaki farkın p değerinin ise 0.47 olduğu görülmektedir, her ikisi de 0.05'ten büyük olduğu için ortalamalar arasında fark olmadığı anlaşılmaktadır. En alttaki tabloda ise gruppardaki yaşın etkisinden arıtlılmış olarak sedimentasyon hızı ortalamaları görülmektedir.

Descriptive Statistics			
Dependent Variable: Sedimentasyon (mm)			
GRUP	Mean	Std. Deviation	N
Kontrol	33.6000	6.6726	30
Tip I DM	29.5333	7.5737	30
Tip II DM	40.9333	7.0707	30
Total	34.6889	8.4848	90

Tests of Between-Subjects Effects					
Dependent Variable: Sedimentasyon (mm)					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5565.203 ^a	3	1855.068	189.453	.000
Intercept	3.910	1	3.910	.399	.529
YAS	3562.447	1	3562.447	363.823	.000
GRUP	8.690	2	4.345	4.44	.643
Error	842.086	86	9.792		
Total	114706.000	90			
Corrected Total	6407.289	89			

a. R Squared = .869 (Adjusted R Squared = .864)

Contrast Results (K Matrix)					
GRUP Simple Contrast ^a	Dependent Variable				
	Sedimentasyon (mm)				
Level 2 vs. Level 1	Std. Error				.840
	Sig.				.727
	95% Confidence Interval for Difference	Lower Bound			-1.375
		Upper Bound			1.964
Level 3 vs. Level 1	Std. Error				.910
	Sig.				.469
	95% Confidence Interval for Difference	Lower Bound			-2.472
		Upper Bound			1.147

a. Reference category = 1

GRUP					
Dependent Variable: Sedimentasyon (mm)					
GRUP	Mean	Std. Error	95% Confidence Interval		
			Lower Bound	Upper Bound	
Kontrol	34.811 ^a	.575	33.669	35.954	
Tip I DM	35.106 ^a	.642	33.830	36.382	
Tip II DM	34.149 ^a	.673	32.811	35.487	

a. Evaluated at covariates appeared in the model: Ya? = 46.17.

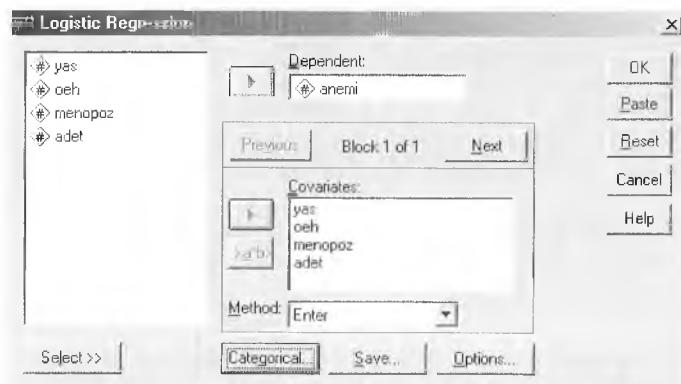
Şekil 9

Örnek 75'teki çalışmada hastalık grubunun bağımsız faktör, yaşın kovaryatı, sedimentasyon hızının bağımlı değişken olduğu kovaryantasyanalizi sonuçları.

Lojistik regresyon analizi

Analyze >> Regression >> Binary Logistic

Yukarıdaki seçimle Logistic Regression diyalog kutusu açılır (Şekil 10). Burada bağımlı değişken *Dependent* kutusuna, prediktörler ya da bağımsız değişkenler *Covariates* kutusuna seçilmelidir. Şekil 10'da Örnek 76'daki çalışmanın analizi için uygun değişken seçiminin yapıldığı görülmektedir.



Şekil 10
SPSS'te "Logistic Regression" diyalog kutusu.

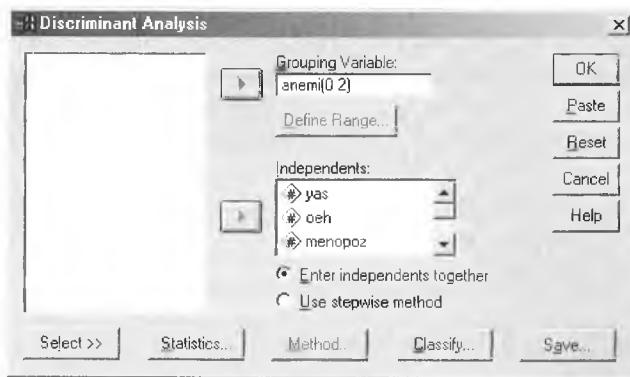
Analize başlamadan önce *Categorical* altmenüsüne girip kategorik değişkenleri *Covariates* listesinden *Categorical Covariates* seçin. *Change Contrast*'tan *Simple*'ı (basit kontrast) ve *Reference Category* olarak da *First*'ü seçin. *Change* butonuna basarak kategorik değişkenlerin kontrast tipini değiştirin.

Diskriminant analiz

Analyze >> Classify >> Discriminant

Yukarıdaki seçimle Discriminant Analysis diyalog kutusu açılır (Şekil 11). Burada bağımlı değişken *Grouping Variable* kutusuna, prediktörler ya da bağımsız değişkenler *Independents* kutusuna seçilmelidir. Şekil 11'de Örnek 77'deki çalışmanın analizi için uygun değişken seçiminin yapıldığı görülmektedir.

Analize başlamadan önce *Classify* altmenüsüne girip *Summary Table* seçeneğini seçerseniz Şekil 6'daki sınıflandırma tablosunu elde edersiniz.

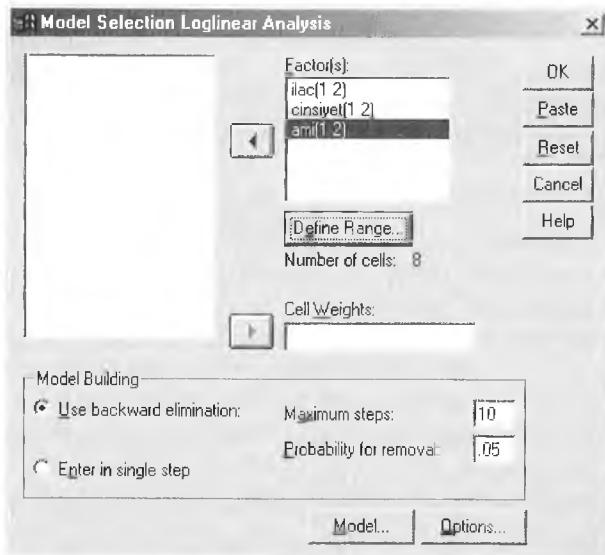


Şekil 11
SPSS'te
“Discriminant
Analysis”
diyalog kutusu.

Log-lineer analiz

Analyze >> Loglinear >> Model Selection

Yukarıdaki seçimle Model Selection Loglinear Analysis diyalog kutusu açılır (Şekil 12). Değişkenler Factor(s) kutusuna seçilmelidir. Şekil 12'de Örnek 78'deki çalışmanın analizi için uygun değişken seçiminin yapıldığı görülmektedir.

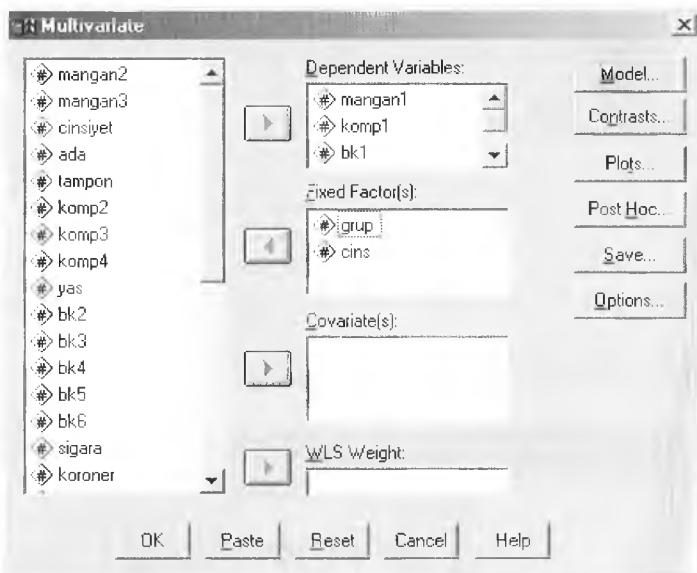


Şekil 12
SPSS'te “Model
Selection Loglinear
Analysis” diyalog kutusu

Çok değişkenli varyans analizi (MANOVA)

Analyze >> General Linear Model >> Multivariate

Yukarıdaki seçimle Multivariate diyalog kutusu açılır (Şekil 13). Burada karşılaştırılacak ortalamaları belirleyen değişkenler, yani bağımlı değişkenler *Dependent Variables* kutusuna, grupları belirleyen değişkenler, yani faktör ya da bağımsız değişkenler *Fixed Factor(s)* kutusuna ve eğer varsa kovaryatlar *Covariate(s)* kutusuna seçilmelidir.



Şekil 13
SPSS'te
"Multivariate"
diyalog kutusu.

Analize başlamadan önce *Contrasts* altmenüsüne girip *Change Contrast*'tan *Simple*'ı (basit kontrast) ve *Reference Category* olarak da *First*'ü seçin. *Change* butonuna basarak grup değişkeninin kontrast tipini değiştirin. Böylece ikişerli karşılaştırmalar 0 ile kodlanmış olan kontrol grubu ile 1 ve 2 ile kodlanmış olan tip I DM ve tip II DM grupları arasında yapılacaktır.

Tanımlayıcı istatistikleri almak için *Options* altmenüsüne girip *Descriptive statistics*'i seçin. Kovaryat etkisinden arıtılmış ortalamaları almak için *Options* altmenüsüne girip *Factor(s) and Factor Interactions* listesinden grup değişkenini *Display Means for* listesine seçin.

Değişkenlerin Şekil 13'deki seçimiyle yapılan analiz sonucu Şekil 7'de görülmektedir.

bölüm 17

TIPTA KARAR VERME

Önemli karar: Tanışal yöntemi uygulayalım mı? Hangisini uygulayalım?

Tanışal yöntemin doğruluk derecesi nedir?

Duyarlılık ve özgürlük

Pozitif ve negatif kestirim değeri

Gelişen teknoloji sonucunda hastalıkların tanı ve tedavisindeki önemli gelişmeler, sağlık alanında tanı ve tedavi yöntemlerinde pahalılığı da birlikte getirmiştir. Bu nedenle günümüzde nasıl tanı konulacağı ya da tedavi edileceğinden başka, tanı ve tedavi yöntemlerinin nasıl en ucuz uygulanabileceğini bilmek ve araştırmak da tıp alanının giderek artan bir ilgiyle üzerinde durduğu bir konu olmaya başlamıştır. Tıbbın konusu insan yaşamı olduğu için, daha az para harcamak kaygısı ile pahalı yöntemlerden tamamen vazgeçmek de olası değildir. Bu nedenle hangi durumlarda hangi yöntemlerin uygulanacağı ile ilgili uygun kararları vermek, doktorların sorumluluk alanını genişletmektedir.

Pahalı teknolojiler nedeniyle, tipta karar vermeye (*medical decision making*) ilişkin sorumluluğun artması, tıp literatüründe bu konuya ilgili istatistik analiz yöntemlerinin kullanımını da uyarmıştır.

ÖNEMLİ KARAR: TANISAL YÖNTEMİ UYGULAYALIM MI? HANGİSİNİ UYGULAYALIM?

Bir hastalığın tanısında kullanılan testi yapıp yapmamakla ilgili karar verirken üzerinde durulması gereken bir kaç nokta önemlidir. Örnek 79'daki hasta için bu noktaları gözden geçirelim.

Elliibeş yaşında pankreas kanseri tanısıyla kemoterapi almakta olan hasta, bir gün önce başlayan, soluk alıp verince artma gösteren yan ağrısı ve nefes darlığı ile başvuruyor. Tanı için pulmoner anjiyografi ve pulmoner ventilasyon-perfüzyon sintigrafisinden birinin kullanımına karar verilecek. Eğer akciğer tromboembolisi tanısı kesinleşirse, intravenöz heparin tedavisine başlanacak. Tedavi edilmediği takdirde hastalık ölümcül olabilir. Bu arada heparinin kanama riskini artırma gibi bir yan etkisi de biliniyor. Hangi tanı yöntemi seçilmelidir?

1. Öykü, fizik inceleme ve diğer laboratuvar testleri sonucunda (henüz test yapılmadan) hastalığın olma olasılığı (prior probability). Bu karar tamamen doktorun klinik deneyimi ile ilgilidir. Hastanın öyküsü akciğer tromboembolisi lehinedir. Fizik incelemenin normal olması tanıyı kesinlikle ekarte etmez. Eğer arteriyel hipoksemisi olduğu da saptanırsa, akciğer trombo-embolisi olasılığı daha da artacaktır.
2. Testin maliyeti. Bu açıdan bakıldığından daha ucuz olan sintigrafi tercih edilebilir.
3. Testin riski. Pulmoner anjiyografide damar içi kateter uygulanması nedeniyle sintigrafiye göre komplikasyon riski daha fazladır.
4. Testin doğruluk derecesi. Sintigrafiye yalancı pozitiflik olabildiği halde, anjiyografide doğruluk derecesi daha yüksektir.
5. Tedavinin yararı ve yan etkileri. Heparin tedavisi oldukça etkili olmakla birlikte kanamayı artırabilir. Bu nedenle yalnız öykü ve fizik incelemeyle yetinip, diğer tanısal işlemleri uygulamadan empirik olarak heparin tedavisi verilirse, belki de yalnızca myaljisi olan bir hastayı gereksiz yere heparinin yan etkileri ile karşı karşıya bırakabilirsiniz.
6. Tedavi edilmediği takdirde hastalığın seyri. Akciğer tromboembolisi tedavi edilmezse, bazı durumlarda ağır seyredebilir. Bu nedenle kesin tanısının konulması ya da ekarte edilmesi için tanısal işlemlerin uygulanması gereklidir.
Yukarıdaki noktalar yalnız her hastalıkta değil, aynı hastalığın birbirinden farklı bulgularıyla başvuran her hastada ayrı olarak değerlendirilir. Bu nedenle doğru karar verilebilmesi için uygulanacak testin hastaya maliyetinden doğruluk derecesine dek bir çok ayrıntıyı bilmek sağlıkla uğraşan kişilere görevi olmalıdır.

TANISAL YÖNTEMİN DOĞRULUK DERECESİ NEDİR?

Duyarlılık ve özgüllük

Genel olarak bir testin doğruluk derecesinin iki yönü vardır. Koroner arter hastalığının tanısında EKG'nin doğruluk derecesini inceleyelim. Önce testin sensitivitesi (duyarlılık: *sensitivity*) denilen kavramı ele alalım. Koroner arter hastalığı olanlarda testin pozitif olması olasılığı (yani EKG'de ST çökmesi olması), testin duyarlılığıdır. Bir testin duyarlılığı ne kadar yüksekse, yalancı-negatiflik olasılığı o kadar azalır. Örneğin 100 koroner arter hastasından 92'sinin EKG'sinde ST çökmesi varsa, EKG'nin koroner arter hastalığı tanısı için duyarlılığı %92'dir; yalancı-negatiflik oranı %8'dir.

$$\text{Duyarlılık} = \text{Gerçek pozitiflik}$$

Bir de testin spesifisitesi (özgüllük: *specificity*) adı verilen başka bir doğruluk ölçütü vardır. Koroner arter hastalığı olmayanlarda testin negatif olması olasılığı (yani EKG'de ST çökmesi olmaması), testin özgüllüğüdür.

Bir testin özgüllüğü ne kadar yüksekse, yalancı-pozitiflik oranı o kadar azalır. Örneğin 100 sağlıklı kişinin 97'sinin EKG'si normal, üçünde EKG'de ST çökmesi varsa, EKG'nin koroner arter hastalığı tanısı için özgüllüğü %97'dir; yalancı-pozitiflik oranı %3'tür.

$$\text{Özgüllük} = 1 - \text{Yalancı pozitiflik}$$

Bir testin duyarlılığı ve özgüllüğü arttıkça, hastalığın tanısındaki doğruluk derecesi artıyor demektir.

Pozitif ve negatif kestirim değeri

ÖRNEK 80

Helicobacter pylori gastritinde üreaz testinin duyarlılık ve özgüllüğünün araştırıldığı bir çalışmada mide biyopsisi ile *Helicobacter pylori* gastriti tanısı konulmuş 100 hasta ile, 100 sağlıklı denek alınmış. Sağlıklı deneklerde *Helicobacter pylori* gastriti olmadığı mide biyopsisi ile gösterilmiştir. Üreaz testinin *Helicobacter pylori* gastriti tanısındaki doğruluk derecesi nedir?

Tablo 1'deki 2X2'lik çapraz tabloyu yaparak işe başlayalım.

Tablo 1 Örnek 80'deki çalışmanın sonuçlarının gösterildiği çapraz tablo.

Üreaz testi	<i>H. pylori</i> gastriti		Toplam
	Var	Yok	
Pozitif	Gerçek Pozitif 94	Yalancı Pozitif 32	126
Negatif	Yalancı Negatif 6	Gerçek Negatif 68	74
Toplam	100	100	200

Yüz *H. pylori* gastriti hastasından 94'ünde, 100 sağlıklı kişiden 68'inde üreaz testi doğru sonuç vermiş. Yani duyarlılık %94, özgüllük %68. Öte yan- dan üreaz testi pozitif olan 126 hastadan 94'ünde (%75) ve üreaz testi negatif olan 74 hastadan 68'inde (%92) test sonuçları doğru çıkmış.

Pozitif testin doğru sonuç verme olasılığına, pozitif testin kestirim değeri ya da pozitif kestirim değeri (*predictive value of positive test, positive predictive value*) adı verilir. Aynı şekilde negatif testin doğru sonuç verme olasılığına, negatif testin kestirim değeri ya da negatif kestirim değeri (*predictive value of negative test, negative predictive value*) adı verilir. O halde üreaz testinin bu çalışma grubu için pozitif kestirim değeri %75, negatif kestirim değeri %92'dir.

Tablo 2'de Tablo 1'den farklı olarak normal denek sayısı 100 değil, 1000'dir. Üreaz testi bu 1000 normal denegenin 320'sinde pozitif, kalan 680'inde negatif çıkacaktır. Bu durumda pozitif kestirim değeri $94/414 = \%23$ ve negatif kestirim değeri $680/686 = \%99$ olacaktır. Örnek 80 ile karşılaştırıldığında, bu iki parametrede çok önemli değişiklikler olmuştur.

Tablo 2 Örnek 80'deki çalışmada 100 yerine 1000 normal denek alındığı zaman, çalışma sonuçlarının gösterildiği çapraz tablo.

Üreaz testi	<i>H. pylori</i> gastriti		Toplam
	Var	Yok	
Pozitif	Gerçek Pozitif 94	Yalancı Pozitif 320	414
Negatif	Yalancı Negatif 6	Gerçek Negatif 680	686
Toplam	100	1000	1100

Bir testin doğruluk derecesini temsil eden bu dört parametreden (yani duyarlılık, özgüllük, pozitif kestirim değeri ve negatif kestirim değeri) duyarlılık ve özgüllük çalışmaya alınan hasta sayısı ne olursa olsun değişmez. Ancak, pozitif ve negatif kestirim değerleri, çalışmaya alınan denek sayısına, daha doğrusu çalışma grubundaki hasta ve normallerin oranına bağlıdır.

Denek sayısı değişikçe bu derece değişen bu parametrelerin yararı nedir? Bu soruyu açıklayabilmek için Örnek 80 ve Örnek 81'deki çalışmaları, çalışma düzeni açısından karşılaştırmak gereklidir. Örnek 80'deki çalışmada kontrollü klinik çalışma düzeni kullanılmıştır. Bu çalışmada hasta ve kontrol gruplarına seçilen deneklerin sayıları tamamen araştırmacı tarafından belirlenmiştir. Çalışmaya alınan hastaların %50'sinin *H. pylori* gastriti hastası olması, araştırmacının seçimi olup, hastalık prevalansından bağımsızdır. Oysa Örnek 81'deki çalışmada iki gruba alınan denek sayıları, araştırmacı tarafından ve belirlenmemiştir. Çünkü bu çalışmada kesitsel çalışma düzeni kullanıldığı için, *H. pylori* gastritinin toplumdaki prevalansı iki gruba girecek denek sayılarını belirlemiştir (Burada "toplum" derken, "mide yakınları olan toplum"u anlamak gereklidir; çünkü çalışmaya alınma kriteri olarak, hastanın mide yakınları ile başvurması koşulu aranmıştır). O halde, mide yakınları olanların $100/1100 = \%9.1$ 'inde *H. pylori* gastriti olmalıdır.

ÖRNEK 81

Örnek 80'deki çalışmaya 100 *H. pylori* gastriti hastası ile 100 sağlıklı denek seçilmiştir. Çalışmaya, seçilmiş deneklerin değil de, Gastroenteroloji polikliniğine mide yakınları ile gelen tüm kişilerin alındığını varsayılmıştır. Polikliniğe mide yakınları ile başvuran 1100 kişiye üreaz testi ve mide biyopsisinde *H. pylori* incelemesi yapılmış ve 100'ünde mide biyopsisi ile *H. pylori* gastriti saptanırken, 1000 kişinin mide biyopsisi normal bulunuyor. Tablo 2'deki verilere dayanarak, üreaz testinin *H. pylori* gastriti tanısındaki pozitif ve negatif kestirim değerleri nelerdir?

Toplumdaki hastalık prevalansını temsil eden oranlarda hasta ve normal deneğin alındığı Örnek 81'deki çalışmadan elde edilen pozitif ve negatif kestirim değerlerinin anlamı nedir?

Mide yakınları ile başvuran bir kişide, üreaz testi yapılmadan önce, *H. pylori* gastriti olması olasılığı $\%9.1$ 'dir.

Üreaz testi yapılır ve sonuç pozitif çıkarsa, hastada *H. pylori* gastriti olma olasılığı $\%23$ 'e çıkar. Yani üreaz testi sonucu pozitif olan dört kişiden yalnız birinde gerçekten *H. pylori* gastriti vardır. Tersine sonuç negatif çıkarsa, *H. pylori* gastriti olasılığı $\%1$ 'e düşer, çünkü negatif kestirim değeri $\%99$ 'dur. O

halde, "üreaz testi, *H. pylori* gastriti tanısı koymaktan çok, hastalığı ekarte etmeye yaramaktadır" denebilir.

Bu kez başka bir soru akla geliyor. Karın ağrısı ile gelen her hastaya üreaz testi yapalım mı? Tabii ki hayır. Tanısal bir yöntemin evrensel bir doğruluk derecesi yoktur. Her hastalık, hatta her hasta için doğruluk derecesi değişir. Bunu aşağıdaki örnekteki iki hasta için inceleyelim.

ÖRNEK 82

Birinci hasta 23 yaşında, bayan. Adet gördüğü zamanlarda karnının sağ üst kadranında kolik tarzında ağrı tanımlıyor. Ağrının yemeklerle ilişkisi yok. **İkinci hasta 65 yaşında, erkek.** Uzun yıllardır aç kaldığında, acılı, ekşili gıdalarda artan epigastrik ağrı tanımlıyor. Geçen yıl yapılan üst endoskopide "mide mikrobu" teşhisini konulduğunu söylüyor. Proton pompa inhibitörü ve antibiyotik kombinasyonu önerilmiş, ama düzenli kullanmadı. Mide yakınmaları son haftalarda artmış. Her iki hastadan da üreaz testi isteniyor. Üreaz testinin *H. pylori* gastriti tanı koymakta yararı var mıdır?

Örnek 82'deki birinci hastada ağrının lokalizasyonu ve ilişkili olduğu etkenler dikkate alındığında *H. pylori* gastriti olma olasılığı çok düşüktür; %1 olduğunu varsayılmış. Buna ön olasılık (*prior probability*) adı verilir (Örnek 81'deki çalışmada ön olasılık, prevalansa eşit olarak varsayılmıştı).

Bu hastada üreaz testi yapılmasının tanımı koymak açısından ne derece yararı olur, inceleyelim. Küsurlu rakamlarla uğraşmamak için bu hasta benzeri 10000 kişide üreaz testi yapıldığını varsayılmış ve aşağıdaki tabloları oluşturalım.

1. Ön olasılık %1 olduğuna göre, mide yakınıması ile başvuran 10000 genç bayan hastadan 100'ünde *H. pylori* gastriti olacak, kalan 9900'ü ise normal olacaktır (Tablo 3).

Tablo 3

Üreaz testi	<i>H. pylori</i> gastriti		Toplam
	Var	Yok	
Pozitif			
Negatif			
Toplam	100	9900	10000

2. Duyarlılık %94 olduğuna göre 100 *H. pylori* gastriti hastasının 94'ünde üreaz testi pozitif çıkacaktır. Geriye kalan 6 kişide ise test yalancı-negatif sonuç verecektir (Tablo 4).

Tablo 4

Üreaz testi	<i>H. pylori</i> gastriti		Toplam
	Var	Yok	
Pozitif	Gerçek Pozitif 94		
Negatif	Yalancı Negatif 6		
Toplam	100	9900	10000

3. Özgüllük %68 olduğuna göre 9900 sağlıklı kişinin %68'inde, yani 6732 sağlıklı kişide üreaz testi negatif, kalan 3168 kişide pozitif çıkacaktır (Tablo 5).

Tablo 5

Üreaz testi	<i>H. pylori</i> gastriti		Toplam
	Var	Yok	
Pozitif	Gerçek Pozitif 94	Yalancı Pozitif 3168	
Negatif	Yalancı Negatif 6	Gerçek Negatif 6732	
Toplam	100	9900	10000

4. Üreaz testinin pozitif çıkacağı kişilerin sayısı toplam olarak $94 + 3168 = 3262$ olacaktır. Aynı şekilde testin negatif çıkacağı kişilerin sayısı toplam olarak $6 + 6732 = 6738$ olacaktır (Tablo 6).

Tablo 6

Üreaz testi	<i>H. pylori</i> gastriti		Toplam
	Var	Yok	
Pozitif	Gerçek Pozitif 94	Yalancı Pozitif 3168	3262
Negatif	Yalancı Negatif 6	Gerçek Negatif 6732	6738
Toplam	100	9900	10000

5. Son olarak üreaz testinin pozitif ve negatif kestirim değerlerini hesaplayalım. Pozitif kestirim değeri $94/3262 = \%3$ 'tür. Negatif kestirim değeri ise $6732/6738 = \%99.9$ 'dur. Yani üreaz testi pozitif çıkarsa, hastanın klinik olarak %1 olduğunu varsayıdığımız *H. pylori* gastriti olasılığı yalnızca %2 artarak, %3'e çıkacaktır. Yani üreaz testi sonucu pozitif çıksa bile, henüz tanı için başka testler gerekecektir. Üreaz testi negatif çıkarsa, önceden %99 olduğunu varsayıdığımız hastanın *H. pylori* gastriti olmama olasılığı %99.9'a çıkacaktır. Bu durumda da üreaz testinin pek yararlı olduğu söylemeyemez.

Şimdi de Örnek 82'deki ikinci hasta için benzer şekilde üreaz testinin yararını inceleyelim. Bu hastanın öyküsü, mide yakınlarının *H. pylori* gastritte bağılı olma olasılığının oldukça yüksek olduğunu düşündürmektedir. Ön olasılığın %90 olduğunu varsayılmı ve yine aynı özelliklerde 10000 kişiye ele alarak üreaz testinin ne derece yararlı olacağını izleyelim ve Tablo 7'yi oluşturalım.

1. Ön olasılık %90 olduğuna göre, mide yakınları ile başvuran benzer özelliklerdeki 10000 hastadan 9000'inde *H. pylori* gastriti olmalı, kalan 1000'i ise normal olmalıdır.
2. Duyarlılık %94 olduğuna göre 9000 *H. pylori* gastriti hastasından 8460'ında üreaz testi pozitif çıkacaktır. Geriye kalan 540 kişide ise test yalancı-negatif sonuç verecektir.
3. Özgüllük %68 olduğuna göre 1000 sağlıklı kişinin %68'inde, yani 680 sağlıklı kişide üreaz testi negatif, kalan 320 kişide pozitif olacaktır.
4. Üreaz testinin pozitif çıkacağı kişilerin sayısı toplam olarak $8460 + 320 = 8780$ olacaktır. Aynı şekilde testin negatif çıkacağı kişilerin sayısı toplam olarak $540 + 680 = 1220$ olacaktır (Tablo 7).

Tablo 7

Üreaz testi	<i>H. pylori</i> gastriti		Toplam
	Var	Yok	
Pozitif	Gerçek Pozitif 8460	Yalancı Pozitif 320	8780
Negatif	Yalancı Negatif 540	Gerçek Negatif 680	1220
Toplam	9000	1000	10000

5. Son olarak üreaz testinin pozitif ve negatif kestirim değerlerini hesaplayalım. Pozitif kestirim değeri $8460/8780 = \%96.4$ 'tür. Negatif kestirim değeri ise $320/1220 = \%27$ dir. Yani üreaz testi pozitif çıkarsa, hastanın kli-

nik olarak %90 olduğunu varsayıdığımız *H. pylori* gastriti tanısı olasılığı %96.4'e çıkacaktır. Klinik olarak zaten yüksek olasılıkla *H. pylori* gastriti olduğunu düşündüğümüz bu hastada üreaz testinin pozitif olması, tanı için pek yarar sağlamamıştır. Üreaz testi negatif çıkarsa, önceden %10 olduğunu varsayıdığımız hastada *H. pylori* gastriti olmama olasılığı %27'ye çıkacaktır.

İki hastada da üreaz testi tanı koymak için işe yaramamıştır. Birinci hastada, yani ön olasılığın düşük olduğu hastada hastalığı ekarte etmek için de işe yaramamışken, ön olasılığın yüksek olduğu ikinci hastada üreaz testinin negatif olması halinde *H. pylori* gastriti olasılığı önemli derecede azalacaktır.

Bu kez de üreaz testinin duyarlılığının %68, özgüllüğünün %94 olduğunu varsayıarak, beyin cimnastiğine devam edelim. Bu durumda Örnek 82'deki iki hastaya talyumlu sintigrafi yapıldığı zaman elde edilecek tablolara göz atalım (Tablo 8 ve 9).

Tablo 8 Örnek 82'deki birinci hastaya benzeyen 10000 hastada üreaz testinin olası sonuçları.

Üreaz testi	<i>H. pylori</i> gastriti		Toplam
	Var	Yok	
Pozitif	Gerçek Pozitif 68	Yalancı Pozitif 594	662
Negatif	Yalancı Negatif 32	Gerçek Negatif 9306	9338
Toplam	100	9900	10000

Tablo 9 Örnek 82'deki ikinci hastaya benzeyen 10000 hastada üreaz testinin olası sonuçları.

Üreaz testi	<i>H. pylori</i> gastriti		Toplam
	Var	Yok	
Pozitif	Gerçek Pozitif 6120	Yalancı Pozitif 60	6180
Negatif	Yalancı Negatif 2880	Gerçek Negatif 940	3820
Toplam	9000	1000	10000

Birinci hasta için pozitif kestirim değeri $68/662 = \%10$, negatif kestirim değeri $9306/9338 = \%99.7$ 'dir. İkinci hasta için pozitif kestirim değeri $6120/6180 = \%99$, negatif kestirim değeri $940/3820 = \%25$ 'tir. Görüldüğü gibi testin özgüllüğü arttıkça, hastalığı ekarte etmek yerine tanı koymak için daha yararlı olmaktadır. Özellikle ön olasılığın yüksek olduğu ikinci hastada üreaz testi pozitifliği ile H. pylori gastriti tanısı $\%99$ olasılıkla konulmaktadır.

- a priori, 153
 alfa hatası, 85
 alternatif hipotez, 84
 analitik istatistik, 83
 analizin gücü, 27
 ANCOVA, 31
 ANOVA, iki yönlü, 31
 ANOVA, tek yönlü, 31
 arınma dönemi, 10
- backward elimination, 244
 bağımlı grup oranları, SPSS'te, 208
 bağımlı iki grup ortalaması, 123
 bağımlı iki grup ortalaması için Student t testi, 125, 127
 bağımlı iki grup ortalaması, güven sınırları, 127
 bağımlı iki grup ortalaması, SPSS'te, 133
 bağımlı oranlar, 200
 bağımlılık, 110
 bağımsız eşzamanlı kontrollü çalışmalar, 8
 bağımsız grup oranları, SPSS'te, 203
 bağımsız iki grup ortalaması, 130
 bağımsız iki grup ortalaması için Student t testi, 132
 bağımsız iki grup ortalaması, güven sınırları, 130
 bağımsız iki grup ortalaması, SPSS'te, 137
 bağımsız oranlar, 179
 bağımsız oranlar, güven sınırları, 180
 bağımsızlık, 110
 bağımsızlık, değişkenler, 111
 bağımsızlık, gruplar, 110
 bağıntı, 214
 bağıntı analizi, 213
 bağıntı analizi, SPSS'te, 222
 bağıntı katsayısı, 219
 bağıntı katsayısı, Pearson, 218
 bağıntı katsayısı, Spearman, 219
 bağıntı, doğrusal olmayan, 217
 basık dağılım, 38, 40
 basit randomizasyon, 20
 basit rastgele örnekleme, 16
 belirteme katsayısı, 222
 beta hatası, 88
 Binary Logistic, SPSS'te, 298
 Bivariate Correlations, SPSS'te, 222
 bloklu randomizasyon, 20
 boksö torbası grafiği, 98, 105
- Charts, SPSS'te, 102
 Classify, SPSS'te, 298
 Cochran Q testi, SPSS'te, 210
 Cochran's Q testi, 201
 Compare Means, SPSS'te, 79, 120, 133, 163
 Cox Regression, SPSS'te, 274
 Cox regresyon analizi, 260
- Crosstabs, SPSS'te, 202, 208
 cv, 66
- çan eğrisi, 38
 çapraz çalışmalar, 10
 çapraz tablolar, 181
 çarpık dağılım, 38, 39, 64
 çeyrek, 74
 çeyreklerarası aralık, 68
 çift-kör, 8
 çok değişkenli analiz yöntemleri, 277
 çok değişkenli analizler, SPSS'te, 295
 çok değişkenli kovaryans analizi, 292
 çok değişkenli varyans analizi, 292
 çok değişkenli varyans analizi (MANOVA), SPSS'te, 300
 çoklu karşılaştırma yöntemleri, 143, 152, 154
- dağılım eğrisi, 37
 dağılım özelliği, 37
 dağılım şekli, 38
 dağılım, basık, 38, 40
 dağılım, çarpık, 38, 39, 64
 dağılım, normal, 38, 119
 dağılım, normal-disi, 119, 128
 dağılımın basıklığı, 38, 69, 74
 dağılımın çarpıklığı, 38, 74
 değer aralığı, 64
 değer etiketi, 54, 58
 değer etiketi, SPSS'te, 53
 değişim katsayısı, 65
 değişken, 33
 değişken adı, 51
 değişken etiketi, 53
 değişken, ikili, 48
 değişken, karakter, 52
 değişken, kesikli sayısal, 35
 değişken, nominal, 48
 değişken, ordinal, 48
 değişken, sansürlü, 250
 değişken, sayısal, 52
 değişken, sürekli sayısal, 35
 değişkenlerin bağımsızlığı, 111
 değişkenlerin dağılım özelliği, 37
 değişkenlerin ölçümü, 33
 deneyel çalışmalar, 7
- Descriptives, SPSS'te, 77, 202, 208
- detrended normal olasılık grafiği, 41
 dış kontrollü çalışmaları, 10
 Discriminant Analysis, SPSS'te, 298
 diskriminant analiz, 31, 286, 288
 diskriminant analiz, SPSS'te, 298
 doğrusal bağıntı, ordinal değişkenler için, 219
 doğrusal bağıntı, sayısal değişkenler için, 218

- doğrusal olmayan bağıntı, 217
 dönüşüm, 119
 dönüşüm, sıra, 120
 Dunnett yöntemi, 156
 duyarlılık, 303
- eş iki grup ortalaması, 123
 etkileşim, 150, 159
 evren, 15, 69
 Explore, SPSS'te, 42, 81, 104
- faktör analizi, 295
 farksızlık hipotezi, 84
 Fisher kesin testi, 187
 forward selection, 243
 Frequencies, SPSS'te, 42, 72, 101
 Friedman iki yönlü varyans analizi, SPSS'te, 174
 Friedman iki-yönlü varyans analizi, 162
- Gehan (genelleştirilmiş wilcoxon) testi, 258
 General Linear Model, SPSS'te, 166, 168, 295, 300
 geometrik ortalama, 63
 gerçek pozitiflik, 303
 gövde-yaprak grafiği, 97
 gözlemlisel çalışmalar, 1
 grafik, boksörtorbası, 98
 grafik, gövde-yaprak, 97
 grafik, kutu-nokta, 98
 grafik, sağlam, 215
 grup varyansları, 132
 grupların bağımsızlığı, 110, 146
 güç, 27, 88
 güç analizi, 91
 güven aralığı, 83, 110
 güven sınırları, 83
 güven sınırları, bağımlı iki grup ortalaması, 127
 güven sınırları, bağımsız iki grup ortalaması, 130
 güven sınırları, bağımsız oranlar, 180
 güven sınırları, tek grup oramı, 178
 güven sınırları, tek grup ortalaması, 116
- H₀ hipotezi, 84
 H₁ hipotezi, 84
 hipotez testi, 84, 110
 hipotez, alternatif, 84
 hipotez, farksızlık, 84
 hipotez, H₀, 84
 hipotez, H₁, 84
 hipotez, iki-yönlü, 88
 hipotez testi, tek grup oramı, 179
 hipotezin yönü, 88
 histogram, 96, 103
- Independent-Samples T Test, SPSS'te, 137
 IQR, 68, 98
 2x2'den büyük çapraz tablolar için ki-kare testi, 190
 iki yönlü ANOVA, 31
 iki yönlü varyans analizi, 149
 iki yönlü varyans analizi, SPSS'te, 166
 ikiden fazla bağımsız grup oranı, 190
 ikili değişken, 34, 48
 ikili normal dağılım, 231
 iki-yönlü hipotez, 88
 iki-yönlü varyans analizi, 145
 istatistiksel önemlilik enflasyonu, 28, 143
- K Independent Samples, SPSS'te, 173
 K Related Samples, SPSS'te, 174, 210
 Kaplan-Meier sağkalım analizi, 256
 Kaplan-Meier, SPSS'te, 267
 kappa istatistiği, 202
 kappa istatistiği, SPSS'te, 209
 karakter değişken, 52
 katsayı, bağıntı, 219
 katsayı, belirtme, 222
 katsayı, değişim, 65
 katsayı, kısmi regresyon, 241
 katsayı, regresyon, 228, 230
 katsayı, varyasyon, 65
 kayıp değer, SPSS'te, 54
 kendi kendine kontrollü çalışmalar, 10
 kesikli sayısal değişken, 35
 kesitsel çalışmalar, 3
 kestirim değeri, negatif, 303
 kestirim değeri, pozitif, 303
 kısmi regresyon katsayısı, 241
 ki-kare, 199
 ki-kare düzeltmesi, 190
 ki-kare testi, 31, 181
 cohort çalışma, 4
 kontrollü çalışma, 7
 kontrollsüz çalışma, 7
 kovaryans analizi, 31, 279
 kovaryans analizi, çok değişkenli, 292
 kovaryans analizi, SPSS'te, 295
 kovaryat, 281
 Kruskal-Wallis analizi, 146
 Kruskal-Wallis tek yönlü varyans analizi, 162
 Kruskal-Wallis tek yönlü varyans analizi, SPSS'te, 173
 kutu-nokta grafiği, 98
 küme analizi, 295
 küme tipi örneklemme, 18
 kümülatif sağkalım olasılığ, 255
- latin kare düzeni, 157
 leptokurtik, 38, 40, 69

- Levene testi, 132
 Life Tables, SPSS'te, 261
 Lilliefors testi, 42, 44
 Linear Regression, SPSS'te, 244
 Loglinear Analysis, SPSS'te, 299
 log-lineer analiz, 31, 289
 log-lineer analiz, SPSS'te, 299
 logrank testi, 259
 lojistik regresyon, 31, 282, 288
 lojistik regresyon analizi, SPSS'te, 298
- MANCOVA, 292
 Mann-Whitney U testi, 133, 162
 Mann-Whitney u testi, SPSS'te, 139
 MANOVA, 292
 Mantel-Haenszel doğrusal ilişki ki-kare testi, 205
 marginal homojenlik testi, 202
 marginal homojenlik testi, SPSS'te, 211
 McNemar testi, 200
 McNemar testi, SPSS'te, 208
 merkezi eğilim ölçütleri, 74
 mod, 63
- negatif kestirim değeri, 303
 Newman-Keuls yöntemi, 155
 nominal değişken, 48
 nominal ölçüm, 34
 Nonparametric Tests, SPSS'te, 136, 173, 208, 210, 211
 nonparametrik testler, 90
 nonparametrik varyans analizi, 162
 normal dağılım, 38, 119
 normal dağılım eğrisi, 37
 normal dağılım testleri, SPSS'te, 42
 normal olasılık grafiği, 41
 normal-dışı dağılım, 119, 128
- odds oranı, 196, 198
 One-Sample T Test, SPSS'te, 120
 One-Way ANOVA, SPSS'te, 163
 oran, 107
 oran, tek grup, 178
 oranlar, bağımlı, 200
 oranlar, bağımlı gruplar, SPSS'te, 208
 oranlar, bağımsız, 179
 oranlar, bağımsız gruplar, SPSS'te, 203
 ordinal değişken, 48
 ordinal değişkenler arasındaki doğrusal bağıntı, 219
 ordinal ölçüm, 34
 ortalama, 62, 107
 ortalama, bağımlı iki grup, 123
 ortalama, bağımlı iki grup, SPSS'te, 133
 ortalama, bağımsız iki grup, 130
 ortalama, bağımsız iki grup, SPSS'te, 137
- ortalama, geometrik, 63
 ortalama sağkalım süresi, 257
 ortanca, 62, 98, 107
- ölçüm düzeyi, 34
 ölçüm tipi, SPSS'te, 55
 ölçüm, değişkenler, 33
 ölçüm, nominal, 34
 ölçüm, ordinal, 34
 ölçüm, sayısal, 35
 önceden planlanmış karşılaştırmalar, 153
 önemli rakam sayısı, 36
 örneklem, 15, 69
 örneklem büyülüğu, 91
 örneklemle, 15
 örneklemle, basit rastgele, 16
 örneklemle, küme tipi, 18
 örneklemle, sistematik, 17
 örneklemle, tabakalı, 17
 özgüllük, 303
- p değeri, 27, 85
 Paired-Samples T Test, SPSS'te, 133
 parametrik testler, 90
 Pearson bağıntı katsayısı, 218
 persentil, 67, 74
 platiyurtik, 38, 40, 69
 posteriori, 154
 post-hoc, 154
 post-hoc testler, 164
 pozitif kestirim değeri, 303
 prospektif çalışmalar, 4
- randomizasyon, 19
 randomizasyon, basit, 20
 randomizasyon, bloklu, 20
 randomizasyon, tabakalı, 20
 randomize blok düzeni, 156
 randomize çalışma, 9
 randomize olmayan çalışma, 9
 Regression, SPSS'te, 298
 regresyon analizi, 227
 regresyon analizi, SPSS'te, 244
 regresyon eşitliği, 228
 regresyon katsayısı, 228, 230
 relativ risk, 194
 Repeated Measures, SPSS'te, 168
 retrospektif çalışmalar, 2
 retrospektif cohort çalışmalar, 6
 rezidüel analizi, 236
 risk fonksiyonu, 257
 risk oranı, 194, 198

- saçılım grafiği, 215
 sağkalım analizi, 249
 sağkalım analizi, SPSS'te, 261
 sağkalım analizinde risk fonksiyonu, 257
 sağkalım olasılığı, 255
 sağkalım süresi, 257
 sansürlü değişken, 250
 sayısal değişken, 52
 sayısal değişkenler arasındaki doğrusal bağıntı, 218
 sayısal ölçüm, 35
 Scheffe yöntemi, 154
 serbestlik derecesi, 86
 Shapiro Wilks' testi, 42, 44
 sıklık tablosu, 73, 76, 100
 sıklık tablosu, SPSS'te, 101
 sınıflandırma tablosu, 285
 sıra dönüşümü, 120
 sistematik örneklemme, 17
 sonradan yapılan karşılaştırmalar, 154
 Spearman bağıntı katsayısı, 219
 standardizasyon, 28
 standart hata, 68, 69
 standart sapma, 64, 69
 stepwise selection, 244
Student t testi, 31
 Student t testi, bağımlı iki grup ortalaması için, 125, 127
 Student t testi, bağımsız iki grup ortalaması için, 132
 Student t testi, SPSS'te, 133, 137
 Survival, SPSS'te, 261
 sürekli sayısal değişken, 35
 sütun formатı, SPSS'te, 55
- t testi**, SPSS'te, 120
t testi, tek grup ortalaması için, 117
 tabakalı analiz, 24
 tabakalı örneklemme, 17
 tabakalı randomizasyon, 20
 tanımlayıcı çalışmalar, 2
 tanımlayıcı grafikler, SPSS'te, 101
 tanımlayıcı istatistik, SPSS'te, 72
 tek grup oranı, 178
 tek grup oranı, güven sınırları, 178
 tek grup oranı, hipotez testi, 179
 tek grup ortalaması için t testi, 117
 tek grup ortalaması testi, SPSS'te, 120
 tek grup ortalaması, güven sınırları, 116
 tek örneklem Kolmogorov-Smirnov testi, 44
 tek yönlü ANOVA, 31, 143
 tek yönlü varyans analizi, 31, 143
 tek yönlü varyans analizi, SPSS'te, 163
 tek-kör, 8
 tekrarlı ölçümeler düzeni, 158
 tekrarlı ölçümeler varyans analizi, SPSS'te, 168
 template, SPSS'te, 55

test istatistiği, 86
tip I hata, 88
tip II hata, 88
tipi, değişken, SPSS'te, 52
2 Related Samples, SPSS'te, 136, 208, 211
Tukey HSD yöntemi, 154

Univariate, SPSS'te, 166, 295
uyum, 202

vaka raporları, 2
varyans, 64
varyans analizi, çok değişkenli, (MANOVA), SPSS'te, 300
varyans analizi, Friedman iki-yönlü, 162
varyans analizi, iki yönlü, 145, 149
varyans analizi, iki yönlü, SPSS'te, 166
varyans analizi, Kruskal-Wallis tek yönlü, 162
varyans analizi, Kruskal-Wallis tek yönlü, SPSS'te, 173
varyans analizi, nonparametrik, 162
varyans analizi, SPSS'te, 163
varyans analizi, tek yönlü, 143
varyans analizi, tek yönlü, SPSS'te, 163
varyans analizi, tekrarlı ölçümler, SPSS'te, 168
varyasyon katsayısı, 65

Wilcoxon işaret testi, SPSS'te, 136
Wilcoxon testi, 128

yalancı pozitiflik, 303
yamılma düzeyi, 85
yaşam tablosu analizi, 253
yaşam tablosu, SPSS'te, 261
Yates düzeltmesi, 190
yayıılma ölçütleri, 64, 74

Z testi, 179, 180

Istatistik analizi (1) uygun yöntemin seçilmesi, (2) yönteme özgü hesaplamaların yapılması ve (3) sonuçların yorumlanması olarak üç aşamada ele alırsak, en kolay aşama hesaplamalardır; çünkü artık güçlü ve hızlı bilgisayarlar ve kolay kullanımı istatistik yazılımlar sayesinde en karmaşık hesaplamalar bile saniyeler içerisinde yapılabiliyor. Uygun formüllerin uygulanmasıyla matematik işlemler yapıldığında ya da bilgisayar klavyesinde birkaç tuşa dokunarak istatistik analizleri yapmak ve ortalaması, oran, p değeri, güven sınırları vb. rakamları hesaplamak çok kolay. Ama istatistik analizde asıl sorun, seçilen yöntemin doğru olup olmadığıdır. Analizin uygun olmayan yöntemlerle yapılması, bilimsel olarak hatalı olmasının yanında, araştırmacıların insan yaşamı ile ilgili sorumlulukları nedeniyle tıp alanında daha da önemlidir. Çünkü bir ilaçın etkinliği, bir tanı yönteminin yararı gibi konularda tıbbi dergi editörlerinin de gözünden kaçabilecek hatalarla yayınlanmış bir makaledeki sonuçlara göre klinik yaklaşımını değiştirecek bir hekim hastalarını riske atmış olacaktır. Bu nedenle sağlık alanında çalışanların hem araştırma yaparken, hem de makale okurken, istatistik analizle ilgili yöntemlere özel önem vermesi kaçınılmazdır.

Bu kitabın amacı, teorik istatistik bilgileri aktarmak değil, medikal istatistik ile ilgili temel kavramları uygulamalı olarak anlatmayaçmaktadır. Böylece üzerinde çalıştığınız araştırmaya en uygun analiz seçenekleri konusunda doğru karar verebilecek, okuduğunuz makalelerdeki sonuçları doğru yorumlayabileceksiniz.

ISBN: 975-6395-47-8



www.pfizer.com.tr