



Ürologlar İstatistik Hakkında Ne Bilmelidir?

Doç. Dr. Fatih TARHAN

Sağlık Bilimleri Üniversitesi

Kartal Dr. Lütfi Kırdar Eğitim ve Araştırma Hastanesi

Uzm. Dr. Meryem Merve ÖREN

Erzurum İl Sağlık Müdürlüğü

Prof. Dr. Ateş KADIOĞLU

İstanbul Üniversitesi İstanbul Tıp Fakültesi

Üroloji Anabilim Dalı, Androloji Bilim Dalı

TÜRK ÜROLOJİ DERNEĞİ

Sunuş

Değerli Meslektaşlarımız ve Tıpta Uzmanlık Öğrencileri,

Türk Üroloji Derneği eğitim, bilim ve teknoloji politikası çerçevesinde TÜAK tarafından koordine edilen kitap dizisinden “Ürologlar İstatistik Hakkında Ne Bilmelidir” kitapçığını üyelerimiz ve tıpta uzmanlık öğrencilerinin kullanımına sunmaktan büyük mutluluk duymaktayız. Tıpta / Ürolojide üretilen bilginin yarılanma süresi beş yıl olup güncel bilginin meslektaşlarımıza ve tıpta uzmanlık öğrencilerine kısa sürede ve evrensel bilgi ışığında ulaştırılması önem kazanmaktadır.

Türk Üroloji Akademisi, **Endoüroloji**'de “Üriner Sistem Taş Hastalığının Tedavisi”, “Robotik Üroloji Güncelleme”, “Ürolojide Lazer Kullanımı”, **Nöroüroloji**'de “Üriner İnkontinans Tanı ve Tedavi”, “Pratik Üroloji El Kitabı”, **Androloji**'de “Erkek ve Kadın Cinsel Sağlığı”, “Tek Karbon Siklusu Üzerinden Etki Eden Tamamlayıcı Tedavilerin Erkek İnfertilitesindeki Yeri”, **Üroonkoloji**'de, “Üroonkoloji El Kitabı”, “Metastatik Böbrek Tümöründe Güncel Tedaviler Kılavuzu”, “Mesane Kanseri Güncelleme”, “Böbrek Kanseri Güncelleme”, “Testis Kanseri”, “Lokal İleri Evre Prostat Kanseri Kılavuzu”, “Mesane Kanseri İntertravezikal Tedaviler Kılavuzu”, “Kastrasyona Dirençli Prostat Kanseri Tedavi Kılavuzu”, “Prostat Kanseri Multiparametrik Manyetik Rezonans Görüntüleme ve Hedefe Yönelik Biopsi Kursu”, **Pediyatrik Üroloji**'de “Çocuk Ürolojisi Güncelleme Kitabı”, **Genel Üroloji**'de “Üretra Darlıklarına Yaklaşım”, “Ürolojide Tıp Hukuku”, “TÜAK / Türkiye ESRU Asistan El Kitabı”, “Ürolojide Perioperatif Süreç Yönetimi”, “Üretra Darlıklarında Tanı, Tedavi ve Takip Protokolleri Kılavuzu”, 2014, 2015, 2016, 2017 ve 2018 EAU Güncelleme Kılavuzu, Güncel Üroloji 2017 ve 2018 kitabını kullanıma sunmuştur.

Genel Üroloji'de “Cerrahi Sanatı El Kitabı”, Üroonkoloji'de “Prostat Kanseri Güncelleme” ve 2018 EAU Güncelleme Kılavuzu eserlerini de en kısa sürede kullanıma sunacaktır.

“Ürologlar İstatistik Hakkında Ne Bilmelidir?” kitapçığı, Doç. Dr. Fatih Tarhan, Uzm. Dr. Meryem Merve Ören ve Dr. Ateş Kadioğlu editörlüğünde hazırlanmıştır. Yayına katkıda bulunan yazarlara teşekkür ederken kitabın meslektaşlarımıza / tıpta uzmanlık öğrencilerine katkısı olan inancımızın tam olduğunu vurgulamak isteriz.

Yayıncılıkta ilk kitapları / dergileri hazırlamak zor; bu yayınları devam ettirmek ise daha da zordur. TÜAK tarafından başlatılan ve koordine edilen bu yayınların elektronik versiyonları da oluşturulmuş ve kullanıma sunulmuştur. Saygılarımızla.

Dr. Ateş Kadioğlu
Türk Üroloji Akademisi Koordinatörü

Dr. R. Cankon Germiyanoglu
Türk Üroloji Derneği Başkanı

ÜROLOGLAR İSTATİSTİK HAKKINDA NE BİLMELİDİR?

Doç. Dr. Fatih Tarhan¹, Uzm. Dr. Meryem Merve Ören², Prof. Dr. Ateş Kadioğlu³

¹Sağlık Bilimleri Üniversitesi Kartal Dr. Lütfi Kırdar Eğitim ve Araştırma Hastanesi

²Erzurum İl Sağlık Müdürlüğü

³Istanbul Üniversitesi İstanbul Tıp Fakültesi Üroloji Anabilim Dalı, Androloji Bilim Dalı

Ürologların tedaviler hakkında bilinçli kararlar almaları, tıbbi literatürü okumaları ve kendi araştırmalarını yürütmeleri için temel bir istatistik bilgisi gereklidir. Bu yazıda bu konu ile ilgili bilgiler özetlenmiştir.

TEMEL KAVRAMLAR

Değişkenler, deneklere ait özelliklerdir. Değişkenler, ikili (örneğin, var veya yok), kategorik (örneğin, zayıf, orta veya şiddetli sınıflandırma grupları, tedavi, plasebo ve kontrol grupları) veya sürekli (akış oranı gibi nesnel olarak ölçülebilir) olarak tanımlanabilir. İkili ve kategorik değişkenler oransal hesaplara ihtiyaç duyarken, sürekli veriler ortamaların test edilmesini gerektirir (1, 2, 3).

“Bağımlı değişken”, değişiklik için ölçülen değişkendir. Bağımlı değişkenin çalışmadaki diğer değişkenlerden etkilendiği düşünülmektedir ve bazen sonuç değişkeni olarak da adlandırılmaktadır. “Bağımsız” değişken, ya araştırmacı tarafından manipüle edilen değişken ya da bağımlı değişkeni etkilediği düşünülen değişken ya da değişkenlerdir. Örneğin, erektil disfonksiyonu (ED) tedavi etmek için yeni bir ilacın araştırıldığı bir çalışmayı düşünelim. Bağımsız değişken grup atamasıdır (yeni tedavi vs. plasebo) ve bağımlı değişken ise erektil fonksiyonun ölçüsüdür. Farklı bir örnek verecek olursak, yaş ve diyabetin idrar akış hızıyla olan ilişkisini araştıran bir çalışmada, yaş ve diyabet bağımsız değişkenler olacak ve idrar akış hızları bağımlı veya sonuç değişkeni olacaktır (4).

Bağımlı ve bağımsız grup kavramları bağımlı ve bağımsız değişken kavramı ile karıştırılabilmektedir. Bağımsız grup, her grupta ayrı deneklerin yer almasıyla mümkündür, bir gruptaki deneklerin aynı zamanda başka gruplarda bulunmaması gerekir. Aslında gruplarda bulunanlar kişiler değil, kişilere ait verilerdir.

Sürekli değişkenlerin nasıl dağılım gösterdiği, kullanılacak istatistik yöntemi seçmek için önemlidir. İnsan popülasyonlarında çalışılan değişkenlerin çoğunun ideal, normal dağılım özelliği gösterdiği varsayılır. Normal dağılımda verilerin, ortalaması, ortanca veya tepe değeri aynı noktada yer alır (1).

Standart sapma (SD) (σ), verilerin ortalamaya göre hangi ölçüde uzakta olduğunu gösterir. Eğer bir sürekli değişken dizisi **normal dağılıyorsa** SD değeri veriler ile ilgili çıkarımlar yapmamıza olanak sağlar;

- Eğer SD küçükse veriler ortalamaya yakın yerlerde, büyükse veriler ortalamadan uzak yerlerde dağılmışlardır.
- Değerlerin %95'i ortalama \pm 2SD aralığında yer alır (1).

Standart hata (SE), örneklem hatasıdır. Standart hata ne kadar küçükse, seçilen örneklemin ana kütleli temsil etme yeteneğinin o kadar yüksek olduğu söylenebilir. Grupların ortalamaları birbirleri ile karşılaştırılırken standart sapma'dan değil, standart hatadan yararlanmak gerekecektir. Örnek ortalamasının 2SE'si %95 güven aralığı içindedir (4).

Güven aralığı (CI), bir parametrenin belirli bir güven seviyesiyle (%95 veya %99), bilinmeyen evren değerini içeren değerler aralığıdır. Güven aralığı, veri aralığını yansıtır ve genellikle yüzde olarak ifade edilir. Diğer bir ifade ile güven aralığı; örneklemin içinden seçildiği evrenden farklı örneklerle aynı hesaplama yapılsa, 100 hesaplama 95'i bu sınırlar içerisinde sonuç verecektir anlamı taşır. Güven aralığının son noktaları güven sınırları olarak adlandırılır. Güven seviyesi, ne kadar emin olabileceğinizi size söyler. Sağlık araştırmalarında çoğunlukla %95 güven seviyesi kullanılmaktadır (3, 5).

Güven aralığını etkileyen en önemli faktörler, örneklemin büyüklüğü ve homojenliğidir. Aralığın parametreyi içermesinin olasılığı, güven seviyesi veya güven katsayısı tarafından belirlenir. Pratikte, %95 CI, sıfır hipotez değerinin üst üste gelmemesi durumunda istatistiksel önemlilik için gösterge olarak kullanılır. Denek sayısı az olduğunda güven aralığı geniş olduğu için ortalama, güven aralığı sınırları içinde kalır ve örneklem için hesaplanan ortalama değerinin evren değerinden istatistiksel olarak farkı bulunmaz (1, 5).

Örnek:

- **Örneklem büyüklüğü 9 iken**, X parametresinin toplum ortalaması 100, örneklem ortalaması 109, güven aralığı 99-119 olduğu bir durumda; toplum ortalaması güven aralığı sınırları içinde kaldığından istatistiksel fark yoktur yorumunu yaparız (4).
- **Örneklem büyüklüğü 100 iken**, X parametresinin toplum ortalaması 100, örneklem ortalaması 109, güven aralığı 107-111 olduğunda ise toplum ortalaması güven aralığı sınırları dışında kaldığından istatistiksel fark vardır yorumunu yapabiliriz (4).

Her ne kadar bir **p** (probability) değeri, sıfır hipotezin reddedilebileceği güvenilirliğini belirlemede ve dolayısıyla gözlemlenen sonucun gücünün belirlenmesinde yardımcı olsa da sonucun kesinliği hakkında bilgi sağlamamaktadır. İstatistiksel anlamlılığın ifadesinde klasik p değeri 0,05 veya %5 olarak kabul edildiğinden, CI %95 güven düzeyinde kullanılır. Genel olarak, eğer gözlenen bir sonuç 0,05 p değerinde istatistiksel olarak anlamlı ise, bu durumda sıfır hipotezi %95 CI içerisinde kalmamalıdır. Güven seviyesi %99 olduğunda p seviyesi $p < 0,01$ olur (5).

Testin gücü, bir testin gerçekte var olan farkı bulabilme yeteneğidir. Güç, örneklem büyüklüğü, etki büyüklüğü ve anlamlılık düzeyi faktörleriyle doğrudan ilişkilidir. Örneklem büyüklüğü araştırmacı için en ulaşılabilir değişkendir. Örneklem büyüklüğünün artırılması, çalışmanın gücünü etkin bir şekilde arttırmaktadır, ancak gücün sadece klinik olarak anlamlı olan farklılıkları tespit etmek için ayarlanması gerekir. Sağlık araştırmalarında gücün %80 ve üzerinde olması beklenir (1, 6).

J. Urol, BJU Int. ve Urology dergilerinin 1982- 2002 yılları arasındaki yayınlanmış makalelerinden 127'si negatif sonuçlu çalışma olarak bulunmuştur. Bunların %65'inde güç analizi eksikliği saptanmıştır. Yazarlar birçok negatif sonuçlu çalışmanın yeterli gücünün olmadığını ve yeterli gücü olmayan çalışmaların yanlış yorumlarla sonuçlandığını bildirmişlerdir (7).

Örneklem büyüklüğü arttıkça, örnek, popülasyonu daha iyi temsil eder hale gelir. Örneklem

hatası, örneklem büyüklüğü arttıkça azalır. Küçük örneklem büyüklüğü daha fazla örnekleme hatası oluşturur, daha geniş güven aralığına sahiptir. Örneklem büyüklüğü artarken güven aralığı da daralır (4, 6). Örneklem büyüklüğü, çalışma tasarımının sadece bir yönüdür. Gerçek hedef yüksek kaliteli bir çalışma tasarlamaktır. **Çalışma, çalışmanın amaçlarına göre yeterli büyüklükte olmalıdır** (6, 8). İkili ve kategorik değişkenler için, örneklem büyüklüğü, temel olayların değerleri, uygun önemlilik seviyesi, çalışmanın istenen gücü ve klinik etki büyüklüğü bilindiği zaman tahmin edilebilir. Sürekli değişkenler için, örnek verilerinin varyansının tahminine de ihtiyaç vardır (1, 6). CONSORT yönergeleri, karşılaştırmalı çalışmalar için iki yönlü test metodolojilerinin kullanılmasını önermektedir (1). Hesaplanan örnek büyüklüğü, kayıplardan dolayı %10-15 oranında artırmaya ihtiyaç duyabilir (1). Örnek büyüklüğünü planlama önemlidir ve neredeyse her zaman zordur. Çalışmaya başlamadan önce bilimsel hedeflerin ortaya çıkarılmasında ve uygun nicel bilginin sağlanmasında dikkatli olunmalıdır. Örneklem büyüklüğü probleminin başarılı çözümü sağlık araştırmaları için, biyoistatistik uzmanlarının ve ilgili diğer uzmanların kapsamlı ve yakın iş birliğini gerektirmektedir (8).

ÖRNEK BÜYÜKLÜĞÜNÜN BELİRLENMESİ VE GÜÇ ANALİZİ NASIL YAPILIR?

Örneklem büyüklüğünü etkileyen faktörler: Veri tipi: Kategorik (yüzde, oran), sayısal (ortalama); Yaygınlık, Alfa (α) önemlilik düzeyi (Tip I hata); Testin gücü ($1 - \beta$: Tip II hata); Etki Büyüklüğü (Δ): Kitlenin büyüklüğü; Çalışma tasarımı (6).

Örneklem büyüklüğünü hesaplamak için gerekli bilgiler: etki büyüklüğü, güç, önemlilik seviyesi, standart sapma, kullanılacak istatistiksel testtir (8).

Önemlilik seviyesi aslında doğru olan sıfır hipotezinin, reddedilmesidir. Tip I hata olarak adlandırılır ve α (alfa) ile gösterilir. Tip I hatası, anlamlılık elde etmek için gereken eşığı tanımlayan p değeriyle ilgilidir. 0,01(%1)- 0,10 (%10) arasında değişebilir sağlık araştırmalarında genellikle 0,05 (%5) olarak alınır (6).

Etki büyüklüğü (Δ) gruplar arasında oluşması tahmin edilen, beklenen farkın yüzde olarak ifadesidir. Araştırmacı önceden yapılmış çalışmalar varsa bunlardan örnek olarak etki büyüklüğü belirleyebilir. Öngörülen etki büyüklüğü küçüldükçe örneklem büyüklüğü artar. Etki büyüklüğünün hesaplanabilmesi için gruplarda ilgilenilen ölçüm bakımından klinik anlamlı bir farklılık oluşması için **ortalamlar arasındaki farkın** en az kaç olduğunun ve **beklenen standart sapma değerinin** tahmin edilmesi gereklidir. Bu değerler daha önceden yapılmış çalışmalardan elde edilebileceği gibi bir pilot çalışmadan da hesaplanabilirler (1, 6, 8).

Yanlış olan sıfır hipotezin hatalı olarak kabul edildiğinde ise bir tip II hatası (β -Beta) oluşur. **Testin gücü** tip II hatadan kaçma gücüdür. Bu ilişki matematiksel olarak Güç = $1 - \beta$ ile ifade edilir. Tip II hata sağlık araştırmalarında en fazla %20 olarak belirlenir. Bu durumda hesaplanan güç %80'dir. Çalışmanın gücü arttıkça örneklem büyüklüğü artar (1, 4).

Güç analizi, verilerin toplanmasından önce veya sonra yapılabilir. Deneysel tasarımda prospektif güç analizi kullanımı genellikle kabul görürken, retrospektif tekniklerin yararı tartışmalıdır (3).

Her istatistiksel yöntemin gücü ve β hatası kendine özgüdür. Bu nedenle örneklem büyüklüğü farklı çalışma düzenleri için farklı formüllerle hesaplanır. İnternette araştırıldığında güç analizi

yazılımı ve çevrimiçi hesaplayıcıları bulunabilir. İlgili kriterlerin yerleştirildiği özel tablolar (<https://www.surveysystem.com/sscalc.htm>; <http://www.stat.ubc.ca/~rollin/stats/ssize/>) veya MedCalc, SAS, Sample Power, NQuery Advisor, PASS, UnifyPow ve Power and Precision gibi hazır paket programlar kullanılabilir. Hazırlanmış grafikler (OC eğrileri-**operating characteristic**), hazır cetveller ya da formüller yardımı ile örnek büyüklüğü saptanabilir. Güç analizi ve örneklem büyüklüğünün saptanması sırasında bir biyoistatistik uzmanına danışılmalıdır.

Örnekler

Örnek 1

Randomize bir çalışma ile stres üriner inkontinans için geleneksel orta üretral sling ile yeni bir teknik karşılaştırılacaktır ve son nokta sosyal kontinanstır. Geleneksel, orta üretral askıların önceki kanıtlara göre %80'lik bir etkinliğe sahip olduğunu ve klinik olarak anlamlı bir farkın %10 olarak tanımlandığını varsayalım. Sağlık araştırmalarında standart olarak kabul gördüğü şekilde p değerini 0,05 ve gücü %80 olarak seçiyoruz. Bu örnek için örnek büyüklüğü, mevcut formüllerden hesaplanabilir. Z değerleri olasılık dağılım tablolarına istinaden bulunur; bu örnekte normal bir dağılım kullanıyoruz. Bu verilerle örnek büyüklüğü her grupta 83 denek olarak hesaplanır. Bununla birlikte, eğer bu yeni tedavinin başarısının orta üretral askıdan daha kötü olabileceğini düşünüyorsak, tek yönlü bir hipotez kurmuş oluruz ve bu doğrultuda hesaplama yapmamız gerekir. Bu durumda, her grupta 108 deneye ihtiyacımız olur. Takipte karşılaşılabileceğimiz kayıpları veya verilerdeki olabilecek kaybı göz önünde bulundurarak %10-%15 oranında ilave denek almayı planlamalıyız. Bu da her bir grupta yaklaşık 118 denek olması demektir. Klinik olarak anlamlı bir fark tanımımızı %20 olarak kabul ettiğimizde hesapladığımız örneklem büyüklüğünün yaklaşık yarısına ihtiyacımız olduğunu görürüz (1).

Örnek 2

Mini-askı (n=75) ile transobturator orta üretral askı'yı (n=55) karşılaştıran bir negatif bulgulu çalışmada, çalışmanın, eğer iki yöntem arasında istatistiksel olarak bir fark var olsaydı bu farkı saptamak için yeterli güce sahip olup olmadığı düşünebilir. Karşılaştırma geriye dönük olarak yapıldığı için örneklem büyüklüğü hesaplaması yapılmamıştır. Bu çalışmayı "post hoc" güç hesaplamasını göstermek için kullanabiliriz. Literatürden transobturator orta üretral askı'nın yaklaşık %85'lik bir etkinliğe sahip olduğunu biliyoruz. Klinik olarak anlamlı bir farkı %10 ve alfa değerini %5 beta değerini ise %80'lik güce karşılık gelen %20 olarak kabul ederek "post hoc" hesaplamamızı yapıyoruz. Bu "post hoc" güç hesaplamasından, çalışmadan geçerli bir sonuç çıkarmak için her bir grupta en az 78 hastaya ihtiyaç duyulduğunu ve bu iki yöntem arasındaki etkinlikteki %10'luk bir farkı saptamak için çalışmanın biraz güçsüz olduğunu söyleyebiliriz (1).

Örnek 3

Sürekli değişkenler için örnekleme istersek;

Benign prostat hipertrofinin tedavisinde yeni bir ilaç araştırıyoruz ve seçilmiş son noktamız akım hızıdır. Çalışma popülasyonumuzun maksimum akış hızı ortalaması 12 ± 4 ml / s'dir. Klinik olarak

anamlı bir iyileşmenin en az 3 ml / sn ile olacağını ve yine p değeri 0,05 ve %80'lik bir güç seçtiğimizi düşünelim. Örnek büyüklüğü hesaplaması ile (yön testi ile tek bir popülasyondan ortalamaları karşılaştırmak için boyut tahmini), grup başına 11 hasta bulunur. Standart sapma 6 ml / s alınırsa, sonucun ölçümü çok daha az hassastır ve her grup için gerekli olan örnek büyüklüğü 25 olur (1).

Örnek 4

Örnek bir çalışmada, prostat kanseri riski artmış erkekler arasında dutasterid ile tedavinin biyopsi ile saptanan prostat kanseri riskini azaltıp azaltmadığı sınınamaktadır. Sıfır hipotezi prostat kanserini azaltmadığıdır. İstatistiksel anlamlılık için p değeri 0,05 ve daha düşük değerler kabul edildi. Bu dutasteridin prostat kanserini azaltmadığı hipotezinin gerçekten doğruyken reddedilmesi olasılığının <%5 olması durumunda, uygun şekilde reddedileceği anlamı taşır. Bu çalışmada, dutasterid kullanan hastalarda plaseboya göre prostat kanseri gelişimi için %22,8 oranında göreceli bir risk azalması olduğu saptanmıştır ($p < 0,001$). (p değeri önceden belirlenen değerlerin çok altındadır). Bu nedenle dutasteridin prostat kanseri riskini bu hasta popülasyonunda %22,8 oranında azalttığına güveniyoruz. Çalışma örneğimizde, sonuç göreceli risk azalma yüzdesi olarak ifade edilmektedir. Sıfır hipotezi dutasteridin prostat kanserini azaltmadığı olduğundan; kabul etseydik relatif risk azaltma yüzdesi %0 olacaktı. Bu çalışmada dutasterid ile tedavi edilen erkeklerde dört yıl boyunca göreceli risk azalması %22,8 (%95 CI %15,2-29,8) idi. Dutasterid ile risk azalmasının sıfır hipotezi, %95 CI değerine düşmez ve bu nedenle sonuç <0,05 olan bir p değerinde istatistiksel olarak anlamlıdır. Ayrıca, %95 CI nispeten dardır. Güven aralığının alt ucunda bile makul derecede büyük bir relatif risk azalması olması (%15,2), sonucun sadece istatistiksel olarak değil, aynı zamanda klinik olarak da anlamlı olduğunu ortaya koymaktadır (5).

DOĞRU İSTATİSTİKSEL TESTİ SEÇME

İstatistiksel test için karar verirken bağımlı değişkenin nasıl ölçüldüğü (sürekli veya kategorik) ve araştırma tasarımı (grup sayısı, bağımlı değişkenin kaç kez ölçüldüğü), çalışmada farklar mı ilişkiler mi inceleniyor) göz önüne alınır. Ayrıca sürekli değişkenler için verilerin normal dağılıma uygunluğu ve varyansların homojenliği test edilir. Bu konular aydınlatıldıktan sonra uygun bir istatistiksel test seçebilir (2).

Normal dağılımı olan veriler ortalama ve SD şeklinde bildirilir. Çarpık veriler için medyan ve çeyrek aralıklar bildirilir. Bazen, çarpık verilerin "normalleştirilmesi" (dönüşüm) istenebilir.

Bağımlı Değişkenlerin Niteliği

Bağımlı değişken, sürekli değişken olarak ölçülürse, ortalama ve standart sapma ve varyans hesaplanır ve parametrik istatistiksel testler kullanılır. Parametrik testler, t-testleri, Analysis of Variance (ANOVA), korelasyon katsayıları ve lineer çoklu regresyondur. Kategorik değişkenler olarak ölçülen değişkenler ise sıklık ya da her kategorideki özne sayısı gibi tanımlayıcı istatistikler üretirler. Bunlar genellikle yüzde olarak bildirilir. Bu durumda non-parametrik testler kullanılır. Bunlar, ki-kare, Fisher's exact test, nokta iki serisel korelasyon katsayıları ve lojistik regresyondur (4).

Çalışma Tasarımı

A. Grup Tasarımları

Grup tasarımları, araştırmacının iki veya daha fazla grubu karşılaştırdığı ve grupların bağımlı değişken açısından farklı olup olmadığını değerlendiren deneylerdir. Randomize kontrollü çalışmalar (RCTs) bu kategoriye girmektedir. Bu tasarımlar çalışmanın önemli unsurları için randomizasyon ve kontrol kullandıklarından, bunlar “deneysel tasarımlar” olarak kabul edilir, bu da araştırmacının bu çalışmalardan neden ve sonuç çıkarabileceği anlamına gelir. “Yarı deneysel tasarımlar” da grup tasarımlarıdır. Bu tür tasarımlar iki veya daha fazla grubu karşılaştırır, ancak çalışmanın önemli öğeleri için randomizasyon veya kontrol kullanmazlar. Retrospektif kohort çalışmaları da iki veya daha fazla grubu karşılaştıran bir grup tasarımı olarak yapılandırılabilir. Bu tür çalışmalar, randomizasyon ve uygun kontrol eksikliği nedeniyle neden ve sonuç veremez. “Tekrarlanan ölçümler tasarımları” grup tasarımları olarak kabul edilir. Tekrarlanan ölçümler tasarımı zaman içinde bir grubu takip eder ve grubu iki veya daha fazla kez değerlendirir (ölçümü tekrarlar). Basit bir örnek, müdahaleden önce bir grubun değerlendirildiği ve aynı grubun müdahaleyi takiben değerlendirildiği önce-sonra tasarımıdır. Buna “kendi deneklerini kontrol olarak kullanan” çalışma da denilir. Aynı bir kontrol grubu olmadığı için tekrarlanan ölçümlerden sebep-sonuç ilişkisi çıkarılamaz (4).

B. İlişkisel Tasarım

Korelasyonel tasarımlar grupları karşılaştırmaz. Korelasyonel tasarımlar, bir grup içindeki iki veya daha fazla değişken arasındaki ilişkiyle ilgilenir. Örneğin yaş ve erektil fonksiyon arasındaki ilişkiyi bulmak. Bu tasarımda grup farklılığı yoktur ve grup içindeki değişkenler arasındaki ilişkiye odaklanılır. Korelasyonel tasarımlar çoklu bağımsız değişkenler ve bir bağımlı değişken arasındaki ilişkiyi açıklayabilir (yaş ve diyabetin erektil fonksiyon ile ne kadar ilgili olduğu gibi) (4).

Dağılımın şekli

Dağılımın normal olup olmadığını anlamak için hipotez testleri (Shapiro-Wilks, D’Agostino-Pearson, Kolmogorov-Smirnov testleri vb.) uygulanabilir. Parametrik testlerin uygulanması sırasında varyansların eşit olup olmadığı F-Testi (Levene Testi) ile kontrol edilir. Bağımsız iki grupta varyansların birbirine oranı 0.5 ile 2 arasında ise kabaca” varyanslar eşittir “denebilir (3, 9).

Tablo 1. Önemlilik testleri

Parametre	Bir grup	İki bağımsız grup	İki eşleştirilmiş grup	>2 bağımsız grup	>2 eşleştirilmiş grup
Sayısal					
Parametrik	Tek grup t	Bağımsız t	Bağımlı t	ANOVA	Tekrarlı
Nonparametrik	Binominal	Mann Whitney	Wilcoxon	Kruskal Wallis	ANOVA
					Friedman
Kategorik	Tek Grup Ki-kare testi	Ki-kare /Fisher	McNemar Sign	Çok gözlü Ki-kare testi	Cochran Q testi

İstatistik Testleri

Parametrik testler: Veriler ölçüm ile elde edilmiş ise, dağılımın normal ve varyansların eşit olduğu durumlarda uygulanır. Bağımlı t-testi, Bağımsız t-testi; Varyans analizi, tekrarlayan ANOVA gibi (3, 9).

Nonparametrik testler: Nitel verilerle dağılımın normal olmadığı nicel veriler için uygulanır. Wilcoxon testi, Mann-Whitney U testi, Kruskal-Wallis Varyans Analizi, Ki-kare testi gibi (3, 9). Bakınız tablo 1.

Bağımlı Değişken Sürekli Olduğunda İstatistik Prosedürleri

A. Gruplar Arasında Fark

Bağımlı değişken bir sürekli değişken olarak değerlendirilirse, tasarım bir grup tasarımıysa ve temel varsayımlar geçerli ise parametrik testler kullanılır. Temel varsayımlar geçerli değilse, (örneğin, veriler normal dağılıma uygun değilse) parametrik olmayan testler uygun olacaktır (Tablo 1).

Bir Grupta Bir Kez Ölçüm Yapılmışsa

Bu durumda tanımlayıcı istatistikler (ortalama, ortanca, oran, standart sapma gibi) verilebilir. Tek grup ortalaması t testi (Nonparametrik karşılığı Binomial Test) ile birden çok değişken arasında ilişki olup olmadığı incelenebilir (3).

Bir Grupta İki Kez Ölçüm Yapılmışsa (İki bağımlı grup)

Bu türdeki çalışmalara önce-sonra türünde ya da kendi-kendine kontrollü çalışmalar adı verilir. Genellikle önceki ve sonraki ölçümler arasında fark olup olmadığı incelenir. Ayrıca iki farklı tanısal yöntemin performansları arasındaki uyumun test edildiği çalışmalar da bu gruba girebilir. Bu türdeki çalışmalarda ise iki yöntem arasında uyum olup olmadığı incelenir. Bağımlı gruplarda t testi (tekrarlanan ölçümler t-testi veya eşleştirilmiş t-testi) uygulanır. Temel varsayımlar geçerli değilse o zaman nonparametrik Wilcoxon testi uygundur (3, 4, 9).

Bir Grupta İkidenden Çok Kez Ölçüm Yapılmışsa (>2 bağımlı grup)

Bu türdeki çalışmalar da önce-sonra çalışmalardır. Ancak ikiden çok kez ölçüm yapılmıştır. Bağımlı gruplarda varyans analizi (Repeated measures ANOVA) uygundur. Temel varsayımlar geçerli değilse nonparametrik Freidman testi uygulanır (3, 4, 9).

İki Grupta Bir Kez Ölçüm Yapılmışsa (İki bağımsız grup)

Bu türdeki çalışmalarda asıl ilgilenilen konu gruplar arasında fark olup olmadığıdır. Bağımsız gruplarda t testi uygulanır. Örneğin, bir araştırmacı ED tedavisi için yeni bir ilacı araştırıyorsa ve iki grup arasında (bir grup yeni tedaviyi alan ve diğer grup plasebo alan) erektil fonksiyonu karşılaştırdıysa, bağımsız gruplarda t testi, doğru istatistiksel test olacaktır. Temel varsayımlar geçerli değilse nonparametrik karşılığı Mann Whitney U testidir (3, 4, 9).

İkiden Çok Grupta Bir Kez Ölçüm Yapılmışsa (>2 bağımsız grup)

Eğer gruplar tek değişken (faktör) ile birbirinden ayrılıyorsa tek yönlü varyans analizi (One way ANOVA) kullanılır. Örneğin, alt üriner sistem semptomları (AÜSS) için yeni bir tedavinin araştırıldığı bir çalışmada üç grup (standart tedavi, yeni tedavi ve plasebo) karşılaştırdığında, bağımsız bir ANOVA testi uygun istatistiksel test olacaktır. Analiz sonucunda fark bulunursa çoklu karşılaştırma testleri (post-hoc yöntemler) ile hangi grupların birbirinden farklı olduğu araştırılır. Post-hoc yöntemler: Tukey HSD, Student-Newman-Keuls, Scheffe, Dunnett, Bonferroni testidir. Eğer dağılım normal değilse, uygun parametrik olmayan test, Kruskal-Wallis tek yönlü varyans analizidir. Kruskal-Wallis testi sonrasında ikili karşılaştırmalar Mann-Whitney U testi ile yapılır. İki veya daha çok değişken (faktör) ile birbirinden ayrılan ikiden çok bağımsız grubu karşılaştırmak için ise iki yönlü varyans analizi (Two way ANOVA) yapılır (3, 4, 9).

İkiden Çok Grupta Birden Çok Kez Ölçüm Yapılmışsa (>2 bağımlı grup)

Bir, iki ya da daha çok grupta en az iki ya da daha çok farklı zamanlardaki ölçümlerin ortalamalarının karşılaştırılması, grup ortalamalarının zaman eğilimlerinin karşılaştırılması için tekrarlı ölçümlü ANOVA yapılır. Örneğin, yeni ilaç, standart tedavi ve plasebo gruplarında tedavi verilmeden önce, tedavi başlangıcından altı hafta ve altı ay sonra deneklerin değerlendirildiği enürezis için yeni bir tedavinin etkinliğini test eden çalışmada tekrarlı ölçümlü ANOVA uygun bir seçimdir. Temel varsayımlar geçerli değilse nonparametrik karşılığı Friedman varyans analizidir (3, 4).

İki ya da daha çok grupta, iki ya da daha çok değişkenin ortalamasının aynı anda karşılaştırılması için ise MANOVA testi uygulanır. AÜSS için farklı üç grubun (yeni tedavi, standart tedavi ve plasebo) karşılaştırıldığı ve değişken olarak uluslararası prostat semptom skorunun (IPSS) alındığı çalışmada, ayrıca diyabetin nasıl bir etki yarattığını da araştırmak istediğimizde, diyabetin varlığı ikinci bir "faktör" olarak eklenir. Uygun istatistiksel test, çok faktörlü bir ANOVA (MANOVA) olacaktır. ANOVA testinden farkı, birden fazla bağımlı değişkenimizin olmasıdır. Bu aynı zamanda, iki veya daha fazla bağımsız değişken ve bir bağımlı değişkenin bulunduğu "çok değişkenli model" in bir örneğidir. MANOVA'nın yararı, ana faktörler arasında bir "etkileşim" olup olmadığını da belirlemesidir. Yeni ilacın AÜSS üzerindeki yararı veya etkisi diyabetin varlığına bağlı olarak farklıysa, etkileşim vardır. Bu testin parametrik olmayan karşılığı PERMANOVA dır (Permutational MANOVA) (3, 4).

B. Değişkenler Arasında İlişki Analizi

Bakınız tablo 2.

Tablo 2. Değişkenler arasında ilişki analizi

Parametre	Regresyon	Korelasyon
Sürekli		
Parametrik	Linear regression	Pearson
Nonparametrik	Nonparametrik lineer regresyon	Spearman
Kategorik	Lojistik regression	Point-Biserial

Korelasyon analizi

Korelasyon, değişkenler arasındaki ilişkiyi ve yönünü ölçmemize yardımcı olur (2). İki değişken arasındaki ilişkiyi değerlendirirken, uygun test genellikle Pearson korelasyon katsayısıdır. En sık bildirilen standart korelasyon budur. Örneğin, yaş ve erektil fonksiyon arasındaki bağlantı veya ilişki araştırılıyorsa, doğru istatistiksel test Pearson korelasyon katsayısı ("r") olacaktır. Pozitif tam ilişki: $r=+1$; Negatif tam ilişki: $r=-1$. Korelasyonun gücü ise $r=0.0-0.24$ zayıf; $r=0.25-0.49$ orta; $r=0.50-0.74$ güçlü; $r=0.75-1.00$ çok güçlü olarak değerlendirilir (2, 3).

Değişkenler arasındaki ilişkinin önemini yorumlamanın bir başka yolu, r karesini hesaplamaktır. Korelasyon katsayısının karesi alındığında, determinasyon katsayısını elde edilir. Katsayı 100 ile çarpıldığında, ikinci değişkeni açıklayan birinci değişkenin yüzdesi bulunur (3, 4).

Pearson korelasyon katsayısından başka bir diğer korelasyon Spearman korelasyon katsayısıdır. Değişkenler sürekli olmadığında ve normal dağılım olmadığında kullanılan non-parametrik bir testtir (3, 4, 9).

Çoklu lineer regresyon

Regresyon analizi, değişkenlerden biri bilindiğinde diğer değişkenin değerinin tahminine yarayan bir analizdir. Regresyon analizi, iki veya daha fazla değişken arasındaki ilişkileri ölçmek için kullanılır. Nedensellik sorgulanmaz. Hem tanımlayıcı hem de çıkarımsal istatistik sağlar. İki değişken-den biri belirli bir birim değiştiğinde diğerinin nasıl bir değişim gösterdiğini inceler (3, 4, 9).

Çoğu zaman, tek sonuçlu değişkenin çoklu değişkenlerle olan ilişkisiyle ilgilenilir. Bu, birden fazla bağımsız değişkenin bir bağımlı değişkenle ilişkilendirildiği "çok değişkenli model" in bir başka örneğidir. Örneğin yaş ve vasküler komorbiditelerin sayısının ve bunların her ikisinin birden erektil fonksiyonla nasıl ilişkili olduğu gibi. Yaşın vasküler komorbiditelerle ilişkili olması nedeniyle, yaşın erektil fonksiyonu etkilemesinin tek nedeni, kişi yaşlandıkça ortaya çıkan vasküler komorbiditelerdeki artıştır. Prosedür her bir değişkeni "saflaştırır". Regresyon, yaşın ve vasküler komorbiditelerin erektil fonksiyon üzerindeki bağımsız katkısını görmemizi sağlar (4).

Araştırmacı, bağımlı değişkeni tahmin eden regresyon modelinde bir dizi bağımsız veya öngörücü değişken kullanabilir. Genel kural, öngörücü değişken başına en az 15 denek olması gerektiğidir. Bu seviyenin altındaki oranlar, yanlış sonuçlara yol açabilir. Tüm regresyon tabanlı modeller, bağımsız / öngörücü değişkenler arasındaki ilişkileri eşzamanlı olarak kontrol ederek çalışırlar. Çoklu doğrusal regresyon ve lojistik regresyon arasındaki fark, bağımlı değişkenin nasıl ölçüldüğüdür. Eğer bağımlı / sonuç değişkeni sürekli ise, doğrusal çoklu regresyon kullanılmalıdır. Bağımlı / sonuç değişkeni kategorik ise lojistik regresyon kullanılmalıdır (4).

Hem multifaktöryel ANOVA ve hem de çoklu regresyonlar çok değişkenli modellerdir. ANOVA istatistik testleri aslında çoklu regresyonun çok spesifik örnekleridir ve ANOVA kullanılarak yapılan analizlerin çoğu çoklu regresyonla da gerçekleştirilebilir. ANOVA analizinden elde edilen sonuçlar genellikle daha sezgiseldir ve bu yüzden ANOVA modelleri genellikle grup tasarımlarından gelen verileri analiz etmek için ilk tercih olarak kullanılması önerilir. Fakat yine de literatürde, ANOVA kullanılması gerektiğini düşündüğünüz yerlerde çoklu regresyon kullanıldığını görebilirsiniz (4).

Çok değişkenli analizlerin uygulanması kapsamlı bir araştırma ve birikim gerektireceğinden doğru sonuca ulaşmak için uzmanlardan destek almak önemlidir.

Bağımlı Değişken Kategorik Olduğunda İstatistik Prosedürleri

A. Gruplar Arasında Fark

Ki-kare ve Fischer Testi

İki gruptaki deneklerin oranını karşılaştırmak ve birbirinden bağımsız olduğunu doğrulamak için kullanılır. Örneğin erektil disfonksiyonun var veya yok olarak değerlendirildiği bir çalışmada, yeni bir ilaç veya placebo tedavisi alan gruplar arasındaki farkı araştırmak için kullanılacak uygun istatistiksel test ki-kare'dir. Bu test, her gruptaki yüzdelerde anlamlı farklılıklar olup olmadığını gösterir. Ki-kare testinin bir kısıtlaması, frekanslar azaldıkça dağılımın bozulmasıdır (4).

Gözlerdeki beklenen frekansların 25'in üzerinde olması durumunda Pearson ki-kare testi uygulanır. Herhangi bir gözdeki beklenen frekansların 25'in altında olması durumunda Yates düzeltmesi uygulanır. Herhangi bir gözdeki beklenen frekans değeri 5'in altında ise Fisher'in kesin ki-kare testinden yararlanılır (2, 4).

Bağımlı değişkenin kategorik olduğu 3'ten fazla gruplarda da Fisher testi veya ki-kare testi uygun testlerdir. Örneğin, üç grup (yeni ilaç grubu, standart tedavi grubu, plasebo grubu) karşılaşıyorsa ve erektil fonksiyonun bağımlı değişkeni ikili olarak ölçülüyorsa (ED var vs. ED yok), uygun testler ki-kare veya Fisher's testi olacaktır. Ancak bağımlı gruplarda tercih edilecek test McNemar (2 ölçüm) testi veya Cochran Q (>2 ölçüm) testi olmalıdır (4).

Lojistik Regresyon

Grup tasarımları değerlendirildiğinde, araştırmacılar gruba bir veya daha fazla bağımsız değişken eklemek isteyebilir. ED için yeni bir ilacın test edildiği ve plasebo grubuna karşı yeni bir ilaç grubunu karşılaştırdığı önceki örnekte olduğu gibi, diyabetin de bağımlı değişkeni etkileyip etkilemediği belirlenmek istenebilir. Bu durumda, tedavi grubunun bağımsız değişkeni (plasebo grubu vs. yeni tedavi grubu) ve ek bağımsız bir diyabet değişkeni (var/yok) olacaktır. Bağımlı değişken kategorik ise (ED var vs. ED yok), lojistik regresyon testi uygun olacaktır. Lojistik regresyon yukarıda anlatılan çoklu regresyon ile benzerlikler gösterir, fakat lojistik regresyonda sonuç değişkeni kategoriktir (4).

B. Değişkenler Arasında İlişki Analizi

Nokta- İki Serili Korelasyon

Bir araştırmacı iki değişken arasındaki ilişkiyle ilgileniyorsa, genel yapı bir korelasyon katsayısıdır. Bağımsız değişken sürekli olduğunda ve bağımlı değişken kategorik olduğunda, bu tür korelasyona nokta-iki serili korelasyon uygulanır. Örneğin erektil fonksiyonun var veya yok olarak değerlendirildiği, erektil fonksiyonun yaş ile ilişkisinin araştırıldığı durumda, uygun test nokta-iki serili korelasyondur. Bu Pearson korelasyonuna benzer ve aynı şekilde yorumlanır. Tek fonksiyonel farklılık, teknik açıdan bakıldığında, korelasyonun nokta-iki serili korelasyon olarak belirtilmesi gerektiğidir (4).

Lojistik Regresyon

Araştırmacı çoğu zaman tek bağımlı değişkenle ilişkili olan çoklu bağımsız değişkenlerin bağlantısıyla ilgilenmektedir. Örneğin, araştırmacı yaş ve komorbiditeler ve bunların her ikisinin birden erektil fonksiyon ile ilişkisini merak ediyor olabilir. Şayet erektil fonksiyon (bağımlı değişken) ikili değişen olarak ölçülürse (ED var vs. ED yok), uygun istatistiksel test lojistik regresyon olacaktır. Regresyon analizi, yaş ve vasküler komorbiditeler arasındaki ilişkiyi kontrol eder ve araştırmacıya her bağımsız / öngörücü değişken ile bağımlı değişken arasındaki bağımsız ilişkiyi gösterir (4).

KAYNAKLAR

1. Hinh P, Canfield SE. Sample size matters: A guide for urologists. *Indian J Urol.* 2011; 27(4): 503–507.
2. Belloio MF, Serrano LA, Stead LG. Understanding statistical tests in the medical literature: which test should I use? *Int J Emerg Med.* 2008 ;1(3):197-9.
3. Özdemir O. *Medikal İstatistik*. 1.baskı. İstanbul: İstanbul Medikal Yayıncılık; 2006
4. Nelson CJ. *Statistics*. AUA University. 2017. https://university.auanet.org/core_topic.cfm?coreid=122.
5. Flechner L, Tseng TY. Understanding results: P-values, confidence intervals, and number need to treat. *Indian J Urol.* 2011; 27(4): 532–535.
6. Sümbüloğlu V, Sümbüloğlu K. *Klinik ve Saha Araştırmalarında Örnekleme Yöntemleri ve Örneklem Büyüklüğü*. 1.baskı. Ankara: Hatipoğlu Yayınları; 2005.
7. Breau RH, Carnat TA, Gaboury I. Inadequate statistical power of negative clinical trials in urological literature. *J Urol.* 2006; 176 (1):263-6.
8. Lenth RV. Some practical guidelines for effective sample size determination. *The American Statistician.* 2001; 55 (3): 187-193.
9. Sümbüloğlu K, Sümbüloğlu V. *Biyoistatistik*. 7.baskı. Ankara: Hatipoğlu Yayınları; 1997.

Türk Üroloji Derneđi/ Turkish Association of Urology

Prof. Nurettin Ökten Sok. 18/2 Şişli / İstanbul

Tel/Phone : +90 212 232 46 89 • Faks/Fax : +90 212 233 98 04

www.uroturk.org.tr