

Deneye Dayalı Çalışmalarda Bazı İstatistiksel Kavramlar

Some Statistical Concepts for Assays: Medical Education

Dr. Nural BEKİROĞLU,^a
Pınar GÜNEL,^b
Dr. Burak AKSU^c

^aBiyostatistik AD,
^cMikrobiyoloji AD,
Marmara Üniversitesi, Tıp Fakültesi,
^bBiyostatistik AD,
Marmara Üniversitesi,
Sağlık Bilimleri Enstitüsü, İstanbul

Geliş Tarihi/Received: 20.10.2008
Kabul Tarihi/Accepted: 15.01.2009

Yazışma Adresi/Correspondence:
Pınar GÜNEL
Marmara Üniversitesi,
Sağlık Bilimleri Enstitüsü,
Biyostatistik AD, İstanbul,
TÜRKİYE/TURKEY
gunelpinar@yahoo.com

ÖZET Laboratuvar çalışmaları, tıp alanında hastalık tanısı, tedavi yönteminin seçimi ve tedavinin izlenmesine yönelik yapılan çalışmaların önemli bir ayağını oluşturmaktadır. İyi bir tıbbi yaklaşım için, laboratuvar çalışmalarının olabildiğince doğru, kesin ve duyarlı bir şekilde yapılması gerekir. Diğer taraftan her laboratuvar, kendi analiz sonuçlarının tutarlı olmasına özen göstermelidir. Bunun için her laboratuvar yöneticisi, kendi laboratuvarına en uygun yöntemi, görüntüleme cihazı veya kit gibi, belirlerken yöntem değerlendirilmesinde kullanılan ilkeler hakkında bilgi sahibi olmalıdır. Yeni bir yöntem seçiminin, yöntem özelliklerinin (uygulama, yöntemsel, performans) ayrı ayrı değerlendirilmesi, pazar araştırması ve önerilen yöntemler arasında uygun yöntem karar verilmesi basamakları vardır. Yöntem seçimini etkileyen başlıca faktörler, yöntemin toplam maliyeti, yöntemin uygulamadaki pratik özellikleri, yöntemin analitik performans özellikleri ve yöntemin tıbbi yararı olarak sayılabilir. Laboratuvarında kullanıma girecek yeni bir yöntem belirlendikten sonra bu yöntemin ölçülebilir özellikleri deneylerle belirlenir. Bu makalede, sağlık alanında çok sık kullanılan ve büyük önem taşıyan in vitro çalışmalarla ilişkili istatistiksel kavramlara değinilmiştir. Bu kapsam içinde, yöntemsel özellikler, deneme özgüllüğü ve duyarlılığı, kestirim değeri gibi performans özellikleri, doğruluk, denemeler arası-deneme içi kesinlik, geri kazanım, doğrusalılık, gözlemciler arası-gözlemler arası güvenilirlik gibi ve laboratuvar yöntemlerinin karşılaştırılmasında kullanılan istatistiksel yaklaşımlar, korelasyon katsayısı, doğrusal regresyon, ki-kare (Mc-Nemar testi), Stuart-Maxwell testi, Wilcoxon testi, t-testi, Mann-Whitney U Testi, F testi, Kruskal-Wallis Testi gibi kavramlar, örneklerle anlatılmaya çalışılmıştır.

Anahtar Kelimeler: Duyarlılık ve özgüllük; sonuçların tekrar üretilebilirliği

ABSTRACT Laboratory studies are major elements in medicine for diagnosis, disease management and follow-up of the treatment response. For a good medical approach, laboratory studies should be correct, precise and sensitive as possible. On the other hand, every laboratory should take care of consistency of their results. When a laboratory director tries to decide for the most appropriate method for his/her laboratory, such as a scanning device or a kit, he/she should have the knowledge about the principles used for the method assessment. Selection of a new method has the following steps to evaluate the characteristics of the method; application, methodological and performance characteristics, market research and decision for an appropriate method through the proposed methods, respectively. The main factors that affect the new method selection are the total cost, the practical features related with application, the analytical performance and the medical benefits. Following the decision of a new method which will be used in the laboratory, the measurable characteristics of the method are determined by some experiments. In this article, some statistical concepts that are frequently used for in vitro laboratory studies are examined. In this frame, methodological characteristics such as assay specificity and assay sensitivity, predictive value, performance characteristics such as accuracy, inter-intra assay precision, recovery, linearity, inter-intra rater reliability, and finally statistical methods such as correlation coefficient, linear regression, chi-square (Mc-Nemar test), Stuart-Maxwell test, Wilcoxon test, t-test, Mann-Whitney U test, F test, Kruskal-Wallis test were explained with examples.

Key Words: Sensitivity and specificity; reproducibility of results

Türkiye Klinikleri J Med Sci 2009;29(4):997-1006

Laboratuvar çalışmaları, tıp alanında hastalık tanısı, tedavi yöntemi seçimi ve tedavinin izlenmesine yönelik yapılan çalışmaların önemli bir ayağını oluşturmaktadır.

İyi bir tıbbi yaklaşım için, laboratuvar çalışmalarının olabildiğince doğru ve duyarlı bir şekilde yapılması gerekir. Diğer taraftan her laboratuvar, kendi analiz sonuçlarının tutarlı olmasına özen göstermelidir. Bunun için her laboratuvar yöneticisi, kendi laboratuvarına en uygun yöntemi, örneğin; görüntüleme cihazı veya kiti belirlerken, yöntem değerlendirilmesinde kullanılan özellikler hakkında bilgi sahibi olmalıdır.^{1,2}

YENİ BİR YÖNTEM SEÇİMİ

Yeni bir yöntemin seçiminde; yöntem özelliklerinin (uygulama, yöntemsel, performans) ayrı ayrı değerlendirilmesi, pazar araştırması ve önerilen yöntemler arasından uygun yönetime karar verilmesi basamakları izlenir. Yöntem seçimini etkileyen başlıca faktörler şunlardır;³

- Yöntemin toplam maliyeti
- Yöntemin uygulamadaki pratik özellikleri
- Yöntemin analitik performans özellikleri
- Yöntemin tıbbi yararı

YÖNTEMİN ANALİTİK PERFORMANS ÖZELLİKLERİ

A. Yöntemsel Özellikler

- A.1. Özgüllük (Specificity)**
 - A.1.1. Tanı testlerinde özgüllük**
 - A.1.2. Analitik özgüllük**
- A.2. Duyarlılık (Sensitivity)**
 - A.2.1. Tanı testlerinde duyarlılık**
 - A.2.2. Analitik duyarlılık**
- A.3. Kestirim Değeri (Predictive Value)**
 - A.3.1. Negative Kestirim Değeri/NKD (Negative Predictive Value/NPV)**
 - A.3.2. Pozitif Kestirim Değeri/PKD (Positive Predictive Value/PPV)**

B. Performans Özellikleri

B.1. Nicel Veriler İçin Performans Özellikleri

B.1.1. Doğruluk (Accuracy)

B.1.2. Kesinlik (Precision)

B.1.2.1. Denemeler arası kesinlik (Inter-assay Precision)

B.1.2.2. Deneme içi kesinlik (Intra-assay Precision)

B.1.3. Geri kazanım (Recovery)

B.1.4. Doğrusallık (Linearity)

B.2. Nitel Veriler İçin Performans Özellikleri

B.2.1. Gözlemler ve Gözlemciler Arası Uyum

B.2.1.1. Gözlemciler Arası Güvenilirlik (Inter-rater Reliability)

B.2.1.2. Gözlemler Arası Güvenilirlik (Intra-rater Reliability)

C. Yöntemlerin Karşılaştırılmasında Kullanılan İstatistikler

C.1. İki Nicel Ölçümlü Yöntemin Karşılaştırılması

C.1.1. Korelasyon Katsayısı

C.1.2. Doğrusal Regresyon

C.2. İki Nitel Ölçümlü Yöntemin Karşılaştırılması

C.2.1. Mc-Nemar (Ki-kare) Testi

C.2.2. Stuart-Maxwell Test

C.2.3. Wilcoxon Test

C.3. Bir Nitel ve Bir Nicel Ölçümlü İki Yöntemin Karşılaştırılması

C.3.1. t-test (Parametrik)

C.3.2. Mann-Whitney U Testi (Nonparametrik)

C.3.3. F Testi (Parametrik)

C.3.4. Kruskal-Wallis Testi (Nonparametrik)

YÖNTEMİN DEĞERLENDİRİLMESİ

Laboratuvarda kullanıma girecek yeni bir yöntem belirlendikten sonra bu yöntemin ölçülebilir özellikleri deneylerle belirlenir.³

■ Yöntem karar kriterlerine (kalite standartlarına, kalite kriterlerine, performans standartlarına-çoğunlukla tıbben müsaade edilen toplam hata değeridir) göre değerlendirilir.

■ Referans yöntemle (altın standart) veya güvenilirliği kanıtlanmış bir yöntemle karşılaştırılır.

■ Yeni yöntemin uygulanıp uygulanmayacağına karar verilir.

A. YÖNTEMSSEL ÖZELLİKLER

A.1. Özgüllük (Specificity)

A.1.1. Tanı testlerinde özgüllük

Tarama/Tanı testlerinde özgüllük, gerçek sağlamlar içinden testin sağlamları belirleyebilme özelliğidir.⁴ Tanı ve tarama testlerinde kullanılan dört gözlü tablo aşağıda Tablo 1’de verilmiştir.

Tablo 1 doğrultusunda; “Gerçek pozitif”, referans ile uygulanan yeni yöntemin birlikte pozitif sonucu verdiği durumdur. “Yalancı pozitif”, referans negatif sonucunu verirken yeni yöntemin pozitif sonucu verdiği durumdur. “Gerçek negatif”, referans ile yeni yöntemin birlikte negatif sonucu verdiği durumdur. “Yalancı negatif”, ise referans pozitif sonucunu verirken yeni yöntemin negatif sonucunu verdiği durumdur.⁵

Düşük “yalancı pozitif” eğilimi gösteren bir yöntem, özgüllüğü yüksek bir yöntem olarak tanımlanır. (Assay Quality Control, <http://www.poultry-health.com/library/serodiss/assayqc.htm>)

Seçicilik (Özgüllük) = $[d/(b+d)]$ (Tablo 1).

A.1.2. Analitik özgüllük

Analitik özgüllük/seçicilik, yöntemin sadece analizi yapılan maddeyi ölçtüğünün bir göstergesidir.

TABLO 1: Tanı / Tarama testlerinde kullanılan dört gözlü tablo.				
	Referans alınan durum		Toplam	
	+	-		
Yeni yöntemin verdiği sonuç	+	a	b	a+b
	-	c	d	c+d
Toplam	a+c	b+d	a+b+c+d	

Analitik özgüllük, yöntemin doğruluğu ile ilgilidir ve yüksek olması istenir. Yöntemin özgüllüğünü belirlemek için interferans (girişim) çalışmaları yapılır.³

A.2. Duyarlılık (Sensitivity)

A.2.1. Tanı testlerinde duyarlılık

Tanı/Tarama testlerinde, testin gerçek hastalar içinden hastaları bulma yeteneğidir. Aşağıda ifade edildiği şekilde hesaplanır.⁴

$$\text{Duyarlılık} = [a/(a+c)] \text{ (Tablo 1).}$$

A.2.2. Analitik duyarlılık

Analitik duyarlılık, analizi yapılan maddenin ölçüm sinyalinde değişiklik meydana getiren en düşük konsantrasyonudur. Bir başka deyişle, farkı gösteren en düşük konsantrasyon miktarıdır. Kalibrasyon eğrisinin eğimi, duyarlılığın göstergesidir (VIM 1993). (VIM: International Vocabulary of Basic and General Terms in Metrology).

Bazı ölçümlerde yüksek duyarlılıktan, düşük özgüllüğe bir değiş tokuş olur. Örnek olarak, kümes hayvancılığında, sürünün problemini belirlemek için kullanılan tarama testinde yüksek duyarlılıkta bir teknik ile çalışılır. Daha sonra, sürüde, pozitif saptanan örneklerle daha özgül testler uygulanır.

Bir testin duyarlılığı, pozitif olduğu bilinen numunelerin bir dizi sulandırımının yapılması (seyreltmeyle) ve reaksiyonun kaybolduğu dilüsyonun (seyreltmeyi) belirlenmesi ile değerlendirilir. (Assay Quality Control, <http://www.poultry-health.com/library/serodiss/assayqc.htm>)

Duyarlılığın alt limiti, bir dizi örneğin tekrarlarına ait ortalamaların ortalamasından 2 standart sapma değeri çıkarılarak hesaplanır. (Pennsylvania State University, Behavioral Endocrinology Lab Technical Notes, Salivary cortisol EIA assay performance characteristics, http://bbh.hhdev.psu.edu/labs/behavioral%20endo%20lab/bel_technical_notesEIA.htm)

Tekrarların ortalamasının ortalaması-(2 x standart sapma)

Örneğin; 6 adet “blank” tüp (0 mg/L) içeren standart bir eğri hazırlandığında ve bu 6 adet tüp ile iki tekrarlı olacak şekilde çalışıldığında, bu denemenin duyarlılığı, her tekrara ait ortalamaların ortalamasından 2 standart sapma değerinin çıkarılması ile elde edilir. (Urinary Iodine Laboratory Methods, <http://www.sph.emory.edu/PAMM/lab/chapter5.htm>)

A.3. Kestirim Değeri (Predictive value)

A.3.1. Negative kestirim değeri/NKD

[(Negative predictive value (NPV))

Tarama/Tanı testlerinde, test negatif (sağlıklı) sonucu verdiği zaman olgunun gerçekten sağlıklı olma olasılığıdır.⁴

$$\text{NKD} = \frac{\text{Gerçek negatif}}{\text{Gerçek negatif} + \text{Yalancı negatif}}$$

$$\text{NKD} = [d/(c+d)] \text{ (Tablo 1).}$$

A.3.2. Pozitif kestirim değeri/PKD

[(Positive predictive value (PPV))

Tarama/Tanı testlerinde test pozitif (hasta) sonucu verdiği zaman, olgunun gerçekten hasta olma olasılığıdır.⁴

$$\text{PKD} = \frac{\text{Gerçek pozitif}}{\text{Gerçek pozitif} + \text{Yalancı pozitif}}$$

$$\text{PKD} = [a/(a+b)] \text{ (Tablo 1).}$$

B. PERFORMANS ÖZELLİKLERİ

B.1. Nicel Veriler için Performans Özellikleri

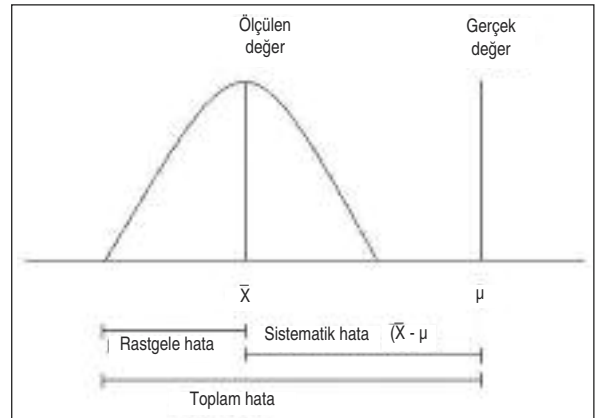
B.1.1. Doğruluk (Accuracy)

Doğruluk, ölçülen değer ile ilgili bir tanımlamadır ve ölçülen değer gerçeğe yakınlığını ifade eder. Toplam hata, rastgele hata ve sistematik hata grafiksel olarak Şekil 1’de verilmiştir.

Analizi yapılan her madde için, bir klinik karar düzeyi ve klinik olarak müsaade edilen toplam hata düzeyi tanımlanmıştır.

$$\% \text{Toplam hata} = \% \text{Sistematik hata} + \% \text{Rastgele hata}$$

Deneylerle elde edilen sistematik hata (SE) ve standart sapmalardan hesaplanan toplam hata düzeyi, klinik olarak müsaade edilen toplam hata (TE) ile karşılaştırılarak performans hakkında karar verilir.



ŞEKİL 1: Bir deneme dağılımındaki hataları gösteren grafik.³

SE, klinik olarak kabul edilen TE’den daha düşük ise, performans yeterlidir.

Klinik olarak müsaade edilen toplam hata düzeyleri için çok sayıda öneri bulunmaktadır. Amerika Birleşik Devletleri’nde yeterlilik testleri sonuçlarından hesaplanarak önerilen değerler [Clinical Laboratory Implementation Amendments 1988 kriterleri] yanında, Avrupa ülkelerinde biyolojik değişkenlik katsayılarını (CV) temel alarak hesaplanan farklı öneriler bulunmaktadır.

B.1.2. Kesinlik (Precision)

Kesinlik, ölçülen değerle ilgili bir tanımlama olarak, farklı koşullar altında tekrarlanan ölçümlerle bulunan değerlerin birbirine yakınlığını ifade eder.

Bir yöntemin üretilebilirliğini (reproducibility) belirlemek için, farklı konsantrasyonlardaki örneklerin yeni yöntemle birçok defa analizi yapılır. Farklı günlerde analizler tekrarlanır. İstatistiksel olarak standart sapmaları ve değişkenlik katsayıları (coefficient of variation) hesaplanır.³

Tablo 2’de laboratuvar içi ve laboratuvarlar arası tekrarlanabilirlik ve üretilebilirliğin karşılaştırılması verilmiştir.

Değişkenlik katsayısı (Coefficient of variation)

Standart sapmanın ortalamaya göre oransal (%) ifadesidir. Standart sapma dağılımının yaygınlığını

TABLO 2: Laboratuvar içi ve laboratuvarlar arası çalışmalarda tekrarlanabilirliğin ve üretilebilirliğin karşılaştırması.

Laboratuvar içi Tekrarlanabilirlik (Repeatability)	Laboratuvarlar arası Üretilebilirlik (Reproducibility)
Aynı Metot	Aynı Metot
Aynı Laboratuvar	Farklı Laboratuvar
Aynı Kişi	Farklı Kişi
Aynı Ekipman	Farklı Ekipman
Farklı Günlerde	

(Tiryaki O, Aysal,P. "Pestisit kalıntı analizlerinde metotların geçerli kılınması," VIII. Ulusal Nükleer Bilimler ve Teknolojileri Kongresi, Türkiye Atom Enerjisi Kurumu-Erciyes Üniversitesi, 15-17 Ekim 2003 Kayseri, Bildiri Özetleri :61.)

gösteren bir ölçümdür. Hassas bir karşılaştırma yapılabilmesi için ölçüyü belirler. Yöntemin ya da ölçümün güvenilirliğini gösterir.⁶

$$\text{Değişkenlik katsayısı (CV)} = \frac{\text{Standart Sapma (SD)}}{\text{Ortalama (M)}} \times 100$$

B.1.2.1. Denemeler arası kesinlik

[(Inter-assay precision, witharound-run (between assays), in a number of assays)]

Denemeler arası kesinlik, uzun dönemli çalışmalarda, aynı yöntemle yapılan farklı deneyler arasındaki yeniden üretilebilirliği belirler. **Bir dizi denemede**, farklı konsantrasyonlardaki numunelerin analiz edilmesi ile değerlendirilen tekrar üretilebilirliktir. Her numune en az iki tekrar (*duplicate*) olacak şekilde çalışılır. (Urinary Iodine Laboratory Methods, <http://www.sph.emory.edu/PAMM/lab/chapter5.htm>)

$$\text{Denemeler arası değişkenlik katsayısı} = \frac{\text{Tekrarların ortalamasının standart sapması}}{\text{Tekrarların genel ortalaması}} \times 100$$

Tablo 3, denemeler arası değişkenlik katsayısı hesabını göstermektedir.

Konsantrasyonları farklı 3 numuneye ait, her deneme kendi içerisinde tekrarları içerecek şekilde 10 farklı deneme yapılır. Sırasıyla farklı numunelere ait denemeler için ikili tekrarların ortalaması alınır. Elde edilen 10 ortalamasının ortalaması ve standart sapması hesaplanır. Bu standart sapmanın, ortalamaya oranının yüzdesi alınarak denemeler

arası değişkenlik katsayısı (inter-assay CV) hesaplanır.

B.1.2.2. Deneme içi kesinlik

(Intra-assay precision, within-run, within assay, in one assay)

Deneme içi kesinlik, bir yöntemde, örnek konumuna bakmaksızın, sonuçların doğru olduğunu gösteren ölçüttür. **Bir denemede**, değişik konsantrasyon seviyelerindeki numunelerin tekrarlarının (replica-

$$\text{Deneme içi değişkenlik katsayısı} = \frac{\text{Tekrarların standart sapmalarının ortalaması}}{\text{Tekrarların genel ortalaması}} \times 100$$

tes) analiz edilmesi ile değerlendirilen tekrar üretilebilirliktir. (Urinary Iodine Laboratory Methods, <http://www.sph.emory.edu/PAMM/lab/chapter5.htm>)

Tablo 4, deneme içi değişkenlik katsayısı hesabını göstermektedir.

Konsantrasyonları farklı 3 numuneye ait birer deneme ele alınarak 15 tekrarı yapılır. Her tekrarın kendi içerisinde ikili tekrarı alınır. Sonuçta bir numuneye ait deneyde 15 x 2= 30 sonuç elde edilir. İkili tekrarların kendi içerisinde ortalaması hesaplanarak standart sapmaları bulunur. Elde edilen 15 standart sapmanın ortalaması ve standart sapması hesaplanır. Bu standart sapmanın, ortalamaya oranının yüzdesi alınarak deneme içi değişkenlik katsayısı (intra-assay CV) hesaplanır.

Not: Interassay/Intraassay CV referans değerleri, üzerinde çalışılan konuya göre Clinical Laboratory Standards Institute'den elde edilebilir. (<http://www.clsi.org/>)

$$\% \text{ Geri kazanım} = \frac{\text{Gözlenen miktar}}{\text{Örneğe eklendiğinde beklenen miktar}} \times 100$$

B.1.3. Geri kazanım (Recovery)

Deneyde, analizi yapılan maddeyi içeren hasta örneğine, bilinen miktarlarda madde içeren çözültiden eklenir, tekrar ölçüm yapılır ve eklenen miktarın yüzde kaçının saptanabildiği hesaplanır.⁷

Örnek ile açıklamak gerekirse; Tablo 5'te farklı konsantrasyonlara sahip iki analite ait geri kazanım sonuçları yer almaktadır.

TABLO 3: Denemeler arası değişkenlik katsayısı hesabı.

Numuneler			
	Numune 1	Numune 2	Numune 3
Denemeler			
1	$T_1 \quad T_2 \Rightarrow \mu_1 = \frac{T_1 + T_2}{2}$	$T_1 \quad T_2 \Rightarrow \mu_1 \text{ (Ort)}$	$T_1 \quad T_2 \Rightarrow \mu_1 \text{ (Ort)}$
2	$T_1 \quad T_2 \Rightarrow \mu_2 = \frac{T_1 + T_2}{2}$	$T_1 \quad T_2 \Rightarrow \mu_2 \text{ (Ort)}$	$T_1 \quad T_2 \Rightarrow \mu_2 \text{ (Ort)}$
3	$T_1 \quad T_2 \Rightarrow \mu_3 = \frac{T_1 + T_2}{2}$	$T_1 \quad T_2 \Rightarrow \mu_3 \text{ (Ort)}$	$T_1 \quad T_2 \Rightarrow \mu_3 \text{ (Ort)}$
4	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
10	$T_1 \quad T_2 \Rightarrow \mu_{10} = \frac{T_1 + T_2}{2}$	$T_1 \quad T_2 \Rightarrow \mu_{10} \text{ (Ort)}$	$T_1 \quad T_2 \Rightarrow \mu_{10} \text{ (Ort)}$
	M: Tekrarların ortalamasının		
	ortalaması = $\sum_{i=1}^{10} \frac{\mu_i}{10}$		
	SD: Tekrarların ortalamasının standart		
	sapması = $\sqrt{\sum_{i=1}^{10} \frac{(\mu_i - M)^2}{10 - 1}}$		
	CV: Değişkenlik katsayısı = $\frac{SD}{M} \times 100$		

T₁: Denemeler için birinci çalışma, T₂: Denemeler için birinci çalışma tekrarı, μ_i: Her bir denemeye ait tekrarların ortalaması, Ort: Ortalama.




Farklı iki numune için eklenen analit çözelti miktarlarının aynı olmasına rağmen (0.54 µg/dL), geri kazanım yüzdesinin değişiklik gösterdiğine dikkat edilmelidir.

Not: Geri kazanım referans değerleri, üzerinde çalışılan konuya göre Clinical Laboratory Standards Institute'den elde edilebilir. (<http://www.clsi.org/>)

B.1.4. Doğrusallık (Linearity)

Değişik konsantrasyonlarda bir dizi standardın en az ikişer kez çalışılıp laboratuvar koşullarında ölçüm ortalamalarının beklenen standart değerlerine göre grafiğe geçirilmesi ve elde edilen grafiğin doğrusal olma halini ifade eder. Doğrusallık, Şekil 2'deki gibi gösterilebilir.

TABLO 4: Deneme içi değişkenlik katsayısı hesabı.

	Tekrarlar (Replicates)	
Numune 1	$T1_1 \quad T1_2 \quad \dots \quad T1_{15}$ $T2_1 \quad T2_2 \quad \dots \quad T2_{15}$  $\mu_1 \quad \mu_2 \quad \dots \quad \mu_{15}$ $\sigma_1 \quad \sigma_2 \quad \dots \quad \sigma_{15}$ (μ_i :Tekerar çiftlerinin ortalaması, σ_i : Tekerar çiftlerinin standart sapması,) $\sigma_i = \sqrt{\frac{\sum_{j=1}^{i-2} (T_{ij} - \mu_i)^2}{2-1}}$	M : İkili tekrarların standart sapmalarının ortalaması = $\sum_{i=1}^{15} \frac{\sigma_i}{15}$ SD : İkili tekrarların standart sapmalarının standart sapması = $\sqrt{\frac{\sum_{i=1}^{15} (\sigma_i - M)^2}{15-1}} \quad \begin{matrix} SD \\ M \end{matrix}$ CV : Değişkenlik katsayısı = $\frac{SD}{M} \times 100$
Numune 2	$T1_1 \quad T1_2 \quad \dots \quad T1_{15}$ $T2_1 \quad T2_2 \quad \dots \quad T2_{15}$  $\mu_1 \quad \mu_2 \quad \dots \quad \mu_{15}$ $\sigma_1 \quad \sigma_2 \quad \dots \quad \sigma_{15}$	M : Tekrarların standart sapmalarının ortalaması SD : Tekrarların standart sapmalarının standart sapması CV : Değişkenlik katsayısı = $\frac{SD}{M} \times 100$
Numune 3	$T1_1 \quad T1_2 \quad \dots \quad T1_{15}$ $T2_1 \quad T2_2 \quad \dots \quad T2_{15}$  $\mu_1 \quad \mu_2 \quad \dots \quad \mu_{15}$ $\sigma_1 \quad \sigma_2 \quad \dots \quad \sigma_{15}$	M : Tekrarların standart sapmalarının ortalaması SD : Tekrarların standart sapmalarının standart sapması CV : Değişkenlik katsayısı = $\frac{SD}{M} \times 100$

T₁: Deneme için birinci çalışma, T₂, T₃, , T₁₅ : Deneme için çalışma tekrarları.

Kalibrasyon Eğrisi

Belli standartlar kullanılarak yapılan bir dizi (en az 5) ölçüm sonuçlarıyla çizilen grafik kalibrasyon eğrisidir. Kalibrasyon eğrisinin doğrusallığı, yöntemin doğrusallığının göstergesidir. Kalibrasyon eğrisi, Şekil 3'teki gibi gösterilebilir.

B.2. NİTEL VERİLER İÇİN PERFORMANS ÖZELLİKLERİ

B.2.1. Gözlemler ve Gözlemciler Arası Uyum (Inter & Intra-rater Reliability)

Sağlık alanında bir uygulayıcı, bir hastalığın varlığı veya yokluğunu ya da normal dışı bir durumun var olup olmadığını göstererek yorumlamak zorundadır. Bu tür bir gözlem, sonucu "evet" veya "hayır"

TABLO 5: Farklı konsantrasyonlara sahip iki analite ait geri kazanım sonuçları.

Numune	Analit Miktarı (ug/dL)	Eklene Analit Çözelti Miktarı (ug/dL)	Beklenen Miktar (ug/dL)	Gözlenen Miktar (ug/dL)	Geri Kazanım (%)
1	.41	.54	.95	.825	%86.8
		.04	.45	.390	%86.7
2	.111	.54	.551	.614	%111.4
		.04	.151	.136	%90.1

(Pennsylvania State University, Behavioral Endocrinology Lab Technical Notes, Salivary cortisol EIA assay performance characteristics, http://bbh.hhdev.psu.edu/labs/behavioral%20endo%20lab/bel_technical_notesEIA.htm)

olan bir çıktıdır, nominal ve aynı zamanda öznel bir ölçüdür.

Genel bir strateji olarak, ölçümlerin güvenilirliğini göstermek için tekrarlar yapılmalıdır ve bu tekrarların birbirleri ile ne kadar uyumlu olduğuna bakılmalıdır. (Inter and intra reliability of a method, <http://homepage.mac.com/aaolmos/KAPPA/kappa.html>)

B.2.1.1. Gözlemler arası uyum

(Intra-rater reliability, Intra-meaning within)

Tek bir gözlemci, aynı konu üzerinde birçok gözlem yaparak gözlem sonuçlarını karşılaştırdığında, gözlemin kabul edilebilirlik derecesi, gözlemler arası uyumu gösterir.

B.2.1.2. Gözlemciler arası uyum

(Inter-rater reliability, Inter-meaning between)

İki veya daha fazla gözlemci aynı konu üzerinde gözlem yaptıklarında, bu gözlemlerin uyumları, gözlemciler arası uyumu gösterir.

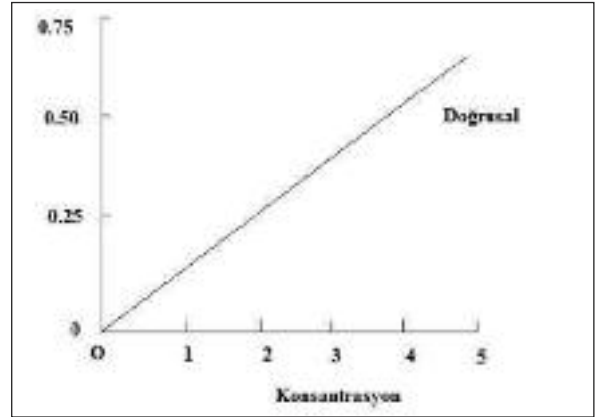
Gözlemciler arası güvenilirliğin ölçülmesi

Farklı gözlemciler tarafından yapılan gözlemlerin sonucu (genellikle ikili dikatomik, örn: evet-hayır, gözlendi-gözlenmedi gibi) adsal/nominal ölçümler olduğunda bu gözlemlerin güvenilirliğini ölçmek için kullanılan yaygın yöntem Kappa istatistiğidir.⁸

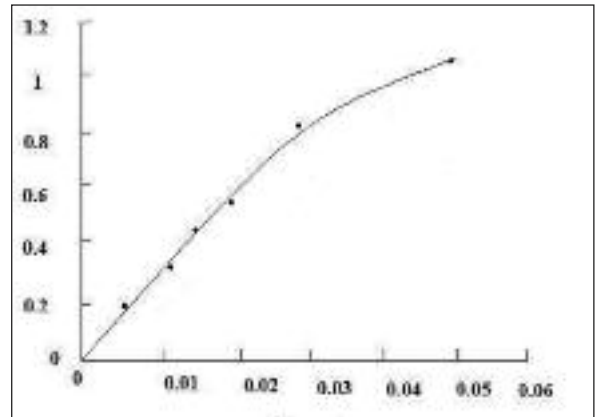
$$K = \frac{(\% \text{Gözlenen uyum}) - (\% \text{Beklenen uyum})}{1 - (\% \text{Beklenen uyum})}$$

Kappa istatistiği (κ)

Kappa istatistiği kategorik verilerde değerlendiriciler arası uyumu ölçen bir katsayıdır. -1 ve +1 ara-



ŞEKİL 2: Doğrusallığın grafik gösterimi.³



ŞEKİL 3: Kalibrasyon eğrisi.³

Not: Doğrusallık için referans değerleri, üzerinde çalışılan konuya göre Clinical Laboratory Standards Institute'den elde edilebilir. (<http://www.clsi.org>).

sında değerler alır. -1 değeri tam uyumsuzluğun, +1 değeri ise mükemmel uyumun göstergesidir. Kappa istatistiği aşağıdaki gibi hesaplanır;

Örnek ile açıklamak gerekirse;

Tablo 6, Landis ve Koch tarafından kullanılan kriterleri göstermektedir.

Tablo 7'de iki radyoloğun aynı hastalar üzerinde yaptığı değerlendirmelerin sonucu yer almaktadır. Bir radyolog diğer radyoloğun değerlendirme sonucunu bilmemektedir.

Tablo 7 doğrultusunda Radyolog A'nın pozitif sonucunu verdiği hasta sayısı 6, olasılığı $6/15=0.40$ ve radyolog B'nin pozitif sonucunu hasta sayısı 9, olasılığı $9/15=0.60$ 'dır. Çarpım kuralından iki radyoloğun pozitif sonuç bakımından aynı fikirde olma olasılıkları $0.40 \times 0.60=0.24$ olmaktadır. Benzer şekilde radyolog A'nın negatif sonucunu verdiği hasta sayısı 9, olasılığı $9/15=0.60$ ve radyolog B'nin negatif sonucunu verdiği hasta sayısı 6, olasılığı $6/15=0.40$ 'tır. Yine çarpım kuralından iki radyoloğun negatif sonuç bakımından aynı fikirde olma olasılıkları $0.60 \times 0.40=0.24$ olur. Böylece her iki radyoloğun değerlendirmelerinde toplamda, $0.24 + 0.24 = 0.48$ olasılık ile hemfikir oldukları söylenebilir. Bu olasılık beklenen uyumun değeridir.

İki radyoloğun birlikte aynı sonucu verdikleri hasta sayıları dikkate alınır, birlikte pozitif sonucu verilen hasta sayısı 3, birlikte negatif sonucu verilen hasta sayısı 3 olmak üzere toplam 6 hastada hemfikir oldukları görülür. Bu durumun olasılığı ise $(3 + 3)/15=0.40$ olacaktır ve bu olasılık gözlenen uyumun değerini gösterecektir. Formül üzerinden yola çıkıldığında aşağıdaki sonuç elde edilecektir.

$$\begin{aligned} \kappa &= \frac{\%40 - \%48}{1 - \%48} \\ &= -0.154 \end{aligned}$$

TABLO 6: Kappa kriterleri.

Kappa indeks	Uyum
< 0.00	Uyuşmazlık
0.00 – 0.20	Zayıf
0.21 – 0.40	Makul
0.41 – 0.60	Orta derecede
0.61 – 0.80	Mükemme yakın
0.81 – 1.00	Mükemmel

(1977, Landis ve Koch tarafından kullanılan kriterler).

Bu sonuç iki değerlendiricinin uyumunun derecesini gösteren ölçüt olmaktadır.

C. YÖNTEMLERİN KARŞILAŞTIRILMASINDA KULLANILAN İSTATİSTİKLER

C.1. İki Nicel Ölçümlü Yöntemin Karşılaştırılması

C.1.1. Korelasyon katsayısı (r)

Korelasyon katsayısı her ikisi de nicel ölçümlü yöntemler arasındaki bağıntının yönünü ve derecesini gösteren bir ölçüdür. "r" ile simgelenir. $-1 \leq r \leq 1$ arasında yer alır. Her iki yönde de 0'dan 1'e doğru yaklaştıkça ilişkinin kuvveti artar. Değişkenler arasındaki ilişki pozitif ise artı işaretlidir. Bunun anlamı değişkenlerin aynı yönde gelişim göstermeleridir. Değişkenlerden birinin değeri arttığında diğeri azalıyor ya da biri azaldığında diğeri artıyor ise ilişki negatif işaretlidir. Bir başka deyişle, değişkenler zıt yönlü gelişim göstermektedirler. Parametrik ve parametrik olmayan yöntemlere göre hesaplanması farklılık gösterir.⁴

C.1.2. Doğrusal regresyon

İki nicel ölçümlü yöntemin arasında doğrusal bir ilişki olup olmadığını göstererek, bu doğrusal ilişkinin bir denklemlerle ifadesini sağlar. Burada sebep sonuç ilişkisi vardır. Korelasyon için sebep sonuç ilişkisi koşul değildir.⁴

TABLO 7: İki radyoloğa ait gözlemler.

		Radyolog B		Radyolog A için Toplam
		Pozitif	Negatif	
Radyolog A	Pozitif	3	3	6
	Negatif	6	3	9
Radyolog B için Toplam		9	6	15

C.2. İki Nitel Ölçümlü Yöntemin Karşılaştırılması

C.2.1. Mc-Nemar (ki-kare) testi

Eşli olgu-denetim çalışmalarında veya aynı toplulukta bir tedavinin veya ilacın öncesi ve sonrası gibi nitel ölçümlerin kıyaslanmasında tablo 2 x 2 (var-yok, pozitif-negatif) ile ifade edilmiş bir ki-kare testidir.^{4,9}

C.2.2. Stuart-Maxwell test

Stuart Maxwell testi, nitel ölçümlü veriler için eşleştirilmiş çiftlerdeki veya tekrarlanmış ölçümlerdeki (örneğin; tedavi öncesi ve tedavi sonrası, iyi ve kötü durumun belirlenmesi) tablo 2 x 2 (dört-gözlü) boyuttaki McNemar testinin özellikle 3 x 3 boyuttaki kontenjans tabloları için olan uzantısıdır.⁴

C.2.3. Wilcoxon test

Parametrik olmayan bir testtir. Eşleştirilmiş dizilerde, bir değişken yönünden aynı bireylerin değişik durumlardaki ölçümleri arasındaki farklılaşmanın değerlendirilmesi amacıyla kullanılır. 3 x 3'ten daha fazla boyuttaki tablo ile ifade edilmiş ölçümlerin kıyaslanmasında kullanılır.^{4,10}

C.3. Bir Nitel ve Bir Nicel ölçümlü İki Yöntemin Karşılaştırılması

C.3.1. t-test (Parametrik)

t-testleri, tek grubu içeren anakitle örnek grubu ortalaması karşılaştırması olabildiği gibi iki anakitleye ait örnek grup ortalamalarının kıyaslanmasında veya eşlendirilmiş dizilerdeki ortalama kıyaslanmasında kullanılır.⁴

C.3.2. Mann-Whitney U Testi (Nonparametrik)

İki bağımsız grubun, ortancalarını (medyan) karşılaştırmak amacıyla kullanılan parametrik olmayan bir testtir. Parametrik test varsayımlarının yerine getirilmediği ve örnek sayısının az olduğu durumlarda kullanılır.^{4,9}

C.3.3. F testi-tek yönlü ANOVA (Parametrik)

İkiden fazla bağımsız grubun, aynı özellik ölçümü açısından ortalamalarının karşılaştırılmasıdır.⁴

C.3.4. Kruskal-Wallis testi (Nonparametrik)

Tek yönlü varyans çözümlemesinin nonparametrik karşılığıdır. İkiden fazla grup söz konusudur ve karşılaştırılan grupların aynı anakitle ortancasına (medyan) sahip olup olmadıklarını test eder.⁴

KAYNAKLAR

1. Arı Z, Aytakin M, Yiğitoğlu R, Uyanık BS, Yar B. [Comparison of analytical performance of different methods used in total CK and CK-MB measuring]. İbni Sina Tıp Der 2002; 7:171-6.
2. Aslan D, Kuralay F, Tanyalcin T, Tanyalcin O. A practical analytical quality report on the effect of carotenemia on neonatal bilirubin levels. Accred Qual Assur 2000;5(2):54-7.
3. Altınışık M. Analitik ölçüm yöntemlerinin laboratuvara kurulması ile ilgili hesaplamalar. ADÜTF Biyokimya AD 2008.
4. Bekiroğlu N. Açıklamalı Biyoistatistik Terimleri Sözlüğü. 1st ed. İstanbul: Nobel Tıp Kitabevi; 1998. p.25-111.
5. Uğurman F, Berktaş MB. [The basic concepts used to evaluate diagnostic tests and risk in medicine]. Türkiye Klinikleri Arch Lung 2003;4 (4):235-40.
6. Türkalp I, Çil Z, Özkazanç D. [Analytical performance of a direct assay for LDL-cholesterol: a comparative assessment versus Friedewald's formula]. Anadolu Kardiyol Derg 2005;5(1):13-7.
7. Güneral F. [An improved radioisotopic assay for the determination of free and esterified carnitine in serum and urine]. Türkiye Klinikleri J Med Sci 1996;16(5):355-9.
8. Dawson B, Trap RG. Research questions about one group. Basic&Clinical Biostatistics. 3rd ed. Singapore: Lange Medical Books/McGraw-Hill; 2001. p.115-17.
9. Altman DG. Comparing groups-categorical data. Practical Statistics for Medical Research. 1st ed. London: Chapman&Hall; 1991. p.265-66.
10. Daniel WW. Procedures that utilize data from two related samples. Applied Nonparametric Statistics. 1st ed. USA: Houghton Mifflin Company; 1978. p.146-51.