Kimyacılar İçin İstatistik ve Kemometri

Prof. Dr. A. Hakan AKTAŞ

PEGEM AKADEMİ

Prof. Dr. A. Hakan AKTAŞ

KİMYACILAR İÇİN İSTATİSTİK VE KEMOMETRİ

ISBN 978-625-7676-61-8
DOI 10.14527/ 9786257676618

Kitap içeriğinin tüm sorumluluğu yazarına aittir.

© 2021, PEGEM AKADEMİ

Bu kitabın basım, yayım ve satış hakları Pegem Akademi Yay. Eğt. Dan. Hizm. Tic. AŞ'ye aittir. Anılan kuruluşun izni alınmadan kitabın tümü ya da bölümleri, kapak tasarımı; mekanik, elektronik, fotokopi, manyetik kayıt ya da başka yöntemlerle çoğaltılamaz, basılamaz, dağıtılamaz. Bu kitap T.C. Kültür ve Turizm Bakanlığı bandrolü ile satılmaktadır. Okuyucularımızın bandrolü olmayan kitaplar hakkında yayınevimize bilgi vermesini ve bandrolsüz yayınları satın almamasını diliyoruz.

Pegem Akademi Yayıncılık, 1998 yılından bugüne uluslararası düzeyde düzenli faaliyet yürüten **uluslararası akademik bir yayınevi**dir. Yayımladığı kitaplar; Yükseköğretim Kurulunca tanınan yükseköğretim kurumlarının kataloglarında yer almaktadır. Dünyadaki en büyük çevrimiçi kamu erişim kataloğu olan **WorldCat** ve ayrıca Türkiye'de kurulan **Turcademy.com** tarafından yayınları taranmaktadır, indekslenmektedir. Aynı alanda farklı yazarlara ait 1000'in üzerinde yayını bulunmaktadır. Pegem Akademi Yayınları ile ilgili detaylı bilgilere http://pegem.net adresinden ulaşılabilmektedir.

1. Baskı: Nisan 2021, Ankara

Yayın-Proje: Şehriban Türlüdür
Dizgi-Grafik Tasarım: Müge Çetin
Kapak Tasarım: Pegem Akademi

Baskı: Ay-bay Kırtasiye İnşaat Gıda Pazarlama ve Ticaret Ltd. Şti.
Çetin Emeç Bulvarı 1314. Cadde No: 37A-B
Çankaya/ANKARA
Tel: (0312) 472 58 55

Yayıncı Sertifika No: 36306
Matbaa Sertifika No: 46661

İletişim

Karanfil 2 Sokak No: 45 Kızılay / ANKARA
Yayınevi: 0312 430 67 50 - 430 67 51
Dağıtım: 0312 434 54 24 - 434 54 08
Hazırlık Kursları: 0312 419 05 60
İnternet: www.pegem.net
E-ileti: pegem@pegem.net
WhatsApp Hattı: 0538 594 92 40

ÖN SÖZ

Kimyacılar için İstatistik ve Kemometri kitabını yazmamdaki amacım, bilim insanları tarafından fiziksel, biyolojik ve sosyal bilimlerde ve beşerî bilimlerde kullanılan çok değişkenli istatistiksel ve matematiksel modelleme prosedürlerine kapsamlı ve okunabilir bir genel bakış sağlamaktır.

Kitap, çok değişkenli istatistiklerin ve matematiksel modelleme tekniklerinin uygun kullanımlarını vurgulamaktadır ve uygulamalı araştırmacıların ve akademisyenlerin bu prosedürleri etkin bir şekilde kullanmalarını sağlayacak uygulamaları öngörür. Her bölüm,

- İstatistik ve kemometrik teknik kullanılarak en iyi yanıtlanabilecek soru türleri,
- İstatistik tekniğin veri gereksinimleri için (örneğin, varsayımlar ve sağlamlık, verilerin özellikleri, ölçüm gereksinimleri ve konu seçimi için gereksinimler) genel bir bakış sağlar,
- Kimyasal analizi gerçekleştirmek için gerekli kararlar alınmasını,
- Sonuçları yorumlarken akılda tutulması gereken hususları göz önüne alır.

Yaygın hatalar gözden geçirilir ve uygulamalı bilimlerdeki endişe verici konuları ele alırken prosedürün uygulanmasını gösteren derinlemesine bir örnek sunar.

Kitabımın bu konuda çalışan bilim insanlarına faydalı olacağını ümit ediyorum.

Prof. Dr. A. Hakan AKTAŞ

Dr. Aktaş, 21.09.1962 yılında İstanbul'da doğdu. Lisans eğitimini Selçuk Üniversitesi Mühendislik-Mimarlık Fakültesi Kimya Mühendisliği programında 1986'da Yüksek lisans eğitimini İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Kimya Mühendisliği programında 1989'da ve Doktora eğitimini de Ankara Üniversitesi Fen Bilimleri Enstitüsü Kimya Programında 1997 yılında tamamladı.

Dr. Aktaş 1986 yılında Akdeniz Üniversitesi Mühendislik Fakültesine Araştırma Görevlisi olarak atandı. 1992 yılında Süleyman Demirel Üniversitesinin kurulmasıyla bu kuruma geçen Dr. Aktaş üniversitenin çeşitli kademelerinde görev aldı. 2007 yılında Doçent, 2013 yılında Profesör unvanı alan Dr. Aktaş halen Fen-Edebiyat Fakültesi Kimya Bölüm Başkanlığı görevini sürdürmektedir.

Dr. Aktaş, Öğretim üyesi olarak sürdürdüğü 1997 yılından beri, 33 yüksek lisans ve 3 doktora öğrencisine danışmanlık yaparak bu öğrencilerin akademik hayatlarına katkı sağlamıştır. Halen yönetiminde 2 doktora ve 3 yüksek lisans öğrencisi çalışmalarına devam etmektedir. Bu süre zarfında Dr. Aktaş 32 adedi SCI tabanlı dergide olmak üzere 59 bilimsel makale yayınlamış, bir adet uluslararası kitap bölümü yazmış ve 103 adet yurt içi ve yurt dışı olmak üzere sempozyumlara bildiri ile katkıda bulunmuştur. H-İndeksi 11 olan Dr. Aktaş'ın yayınladığı makalelerine 350'nin üzerinde atıf' ı bulunmaktadır.

2004 yılından beri Kemometri ile ilgili çeşitli uluslararası dergilerde yayınlar yapan Dr. Aktaş, bu çalışmalarını 2016 yılında Kemometrinin Dünya çapında en iyi dergilerinden biri olan "Journal of Chemometrics" dergisinde yaptığı yayınla taçlandırmıştır. Dr. Aktaş çalışmalarına halen devam etmektedir.

Dr. Aktaş evli ve 2 erkek çocuk babası olarak yaşamına devam etmektedir.

ORCID No: 0000-0003-2327-4031

İÇİNDEKİLER

Ön Söz iii
Prof. Dr. A. Hakan AKTAŞ v

1. BÖLÜM
İSTATİSTİK KEMOMETRİYE GİRİŞ

1.1 İstatistik nedir? 1
1.2 Analitik Metotların Sınırlandırılması 1
1.3 Hataların Sınıflandırılması 2
1.4 Kesinlik ve Doğruluk 4
1.5 Sistematik Hatalar Nasıl Azaltılabilir? 5

2. BÖLÜM
ANALİZ SONUÇLARININ VERİLMESİNDE İSTATİSTİKSEL KAVRAMLAR

2.1. Genel İstatistik Kavramlar 9
2.2. Güven Aralığı 12
2.3. Rastgele Hataların Dağılımı 14
2.4. Normal Dağılım Eğrisinden Yararlanma Şekilleri 16

3. BÖLÜM
DENEYSEL VERİLERİN DEĞERLENDİRİLMESİ

3.1. Anlamlı Testler 19
3.1.1. Bilinen bir değer ile deneysel bir ortalamanın karşılaştırılması 19
3.2. İki Örneğin Ortalamalarının Karşılaştırılması 20
3.3. Anlamlı Testlerde z ve t Testlerinin Uygulanması 22
3.3.1. Popülasyon ortalaması (μ) ve örneklem ortalaması ($\bar{x}$) 22
3.3.2. t testi 22
3.3.3. z testi 24
3.3.4. z- skoru ile sistematik hata tespiti 25
3.3.5. F testi 25
3.4. Atılacak Değeri Tespit Etme 27
3.4.1. Grubbs testi 28
3.4.2. Q testi (Dixon testi) 29

3.5. Varyans Analizi.....31
3.5.1 ANOVA kavramları ve tek faktörlü ANOVA.....31

4. BÖLÜM
KORELASYON VE REGRESYON

4.1. Korelasyon ve Regresyon.....37
4.1.1.Korelasyon katsayısı.....38
4.2. Doğrusal Regresyon.....39
4.3. Regresyon Doğrusunun Eğim ve Kesim Noktasındaki Hatalar.....40
4.4. Derişimin Rastgele Hatasının Hesaplanması 42
4.5. Standart Ekleme Yöntemi.....43

5. BÖLÜM
KEMOMETRİK YÖNTEMLERE GİRİŞ

5.1. Deneysel Tasarım 47
5.2. Varyansın İki-yönlü Analizi.....49
5.3. Kemometri ve Deneysel Tasarım 52
5.4. Faktöriyel Tasarım.....52

6. BÖLÜM
KEMOMETRİDE KULLANILAN KALİBRASYON YÖNTEMLERİ

6.1. Giriş.....57
6.2. Matris Araçları ve İşlemleri 58
6.2.1. Veri matrisi 58
6.2.2. Basit matris işlemleri 61
6.2.3. Matris çarpımı.....62
6.2.4. Karelerin toplanması 64
6.2.5. Matrisin tersi (inversi).....66
6.2.6. Eş zamanlı denklemler.....68
6.2.7. Kuadratik formlar 68
6.3. Çok Değişkenli Bileşen Analizi 69
6.3.1. Değişkenlerin doğrusal bileşenleri 70
6.4. Temel Bileşen Analizi (PCA) 71
6.5. Tek Değişkenli Kalibrasyon 76

6.5.1. Klasik tek değişkenli kalibrasyon yöntemi 77
6.5.2. Ters tek değişkenli kalibrasyon yöntemi 80
6.6. Çok Değişkenli Kalibrasyon Yöntemleri 82
6.6.1. Klasik en küçük kareler kalibrasyonu (CLS) 82
6.6.2. Ters en küçük kareler kalibrasyonu (ILS) 85
6.6.4. Temel bileşen regresyonu (PCR).....92
6.6.5. Kısmi en küçük kareler regresyonu (PLS) 105
6.6.6. Bağımsız bileşen analizi (Independent component analysis (ICA)).....109
6.6.7. Yapay sinir ağları yöntemi (Artificial neural networks, ANN).....111

KAYNAKÇA 119

1. BÖLÜM

İSTATİSTİK KEMOMETRİYE GİRİŞ

1.1 İstatistik nedir?

İstatistik veya **sayıtım**, belirli bir amaç için veri toplama, tablo ve grafiklerle özetleme, sonuçları yorumlama, sonuçların güven derecelerini açıklama, örneklerden elde edilen sonuçları kitle için genelleme, özellikler arasındaki ilişkiyi araştırma, çeşitli konularda geleceğe ilişkin tahmin yapma, deney düzenleme ve gözlem ilkelerini kapsayan bir bilimdir. Belirli bir amaç için verilerin toplanması, sınıflandırılması, çözümlenmesi ve sonuçlarının yorumlanması esasına dayanır.

İstatistiği öğrenmedeki amaç, bir araştırmada elde edilen verilerin uygun istatistiksel yöntemler kullanılarak yorumlanacağını bilmektir.

İstatistiksel yöntemler, toplanmış verilerin özetlenmesi veya açıklanması amacıyla kullanılır. Bu tür bir yaklaşım betimsel istatistik adını alır. Buna ek olarak verilerdeki örtüşmelerin (kalıplar veya örüntüler), gözlemlerdeki rassallığı ve belirsizliği göze alacak şekilde, üzerinde çalışılan ana kütle veya süreç hakkında sonuç çıkarma amacıyla modellenmesi, çıkarımsal istatistik adını alır. Hem betimsel istatistik hem de tahminsel istatistik, **uygulamalı istatistiğin** parçaları olarak sayılabilir. Matematiksel istatistik adı verilen disiplin ise konunun teorik matematiksel altyapısını inceleyen disiplindir.

1.2 Analitik Metotların Sınırlandırılması

Analizcinin görevi; kullanılan analitik işlemlerin doğru uygulanması ve mümkün olduğu kadar gerçek değere yakın bir sonuç elde etmektir. Analizi yapanların yapılan çalışmada kesinliğin, hatanın kaynağının ve kullanılan metodun doğruluğu hakkında yeterli bilgisi olmaması, sonuçların beğenilirliği ve güven düzeyini çok azaltacaktır. Kantitatif analiz sadece bir örnek olarak, tek bir kararlılık taşıma ve alınan değeri reddedilemez olduğunu iddia eden bir olgu değildir. Kantitatif analiz basitçe, bir numune alarak, tek bir belirleme yapılması ve daha sonra elde edilen değerin çürütülemez olduğunu iddia eden bir olgu değildir. Aynı

zamanda diğer iyonlar, elementler ve bileşiklerin muhtemel etkilerinin yanı sıra değerlerinin istatistiksel dağılımının kimya bilgisini de gerektirir [1].

Bu bölümün amacı aşağıda verildiği gibi özetlenebilir.

1. Kullanılan bazı terimlerin açıklanması ve analitik sonuçların uygulanmasında klasik istatistik işlemlerin tanımlanması
2. Ölçümün optimizasyon işlemlerinin ve bunların dizaynları için metot bilgilerinin verilmesi
3. Kemometrinin, bilginin analitik veriden elde ettiğini arttırmayı sağlayabildiği alanın ana hatlarını çizmek

1.3 Hataların Sınıflandırılması

Her analizde veya ölçmede bir hata vardır. Hatasın hiçbir analiz veya ölçme yapılamaz. Analizde yapılan hatalar genel olarak ikiye ayrılır.

1. Rastgele (önlenemeyen) hatalar
2. Sistematik (önlenebilen) hatalar (Biaslar)

Bir sonucun hatası, bu iki hatanın toplamı kadardır. Ortalama değerin hatası da bir sonucun hatası gibidir.

$$E = E_T + E_S \tag{1.1}$$

bağıntısı ile verilir. Burada E_T rastgele, E_S ise sistematik hatadır.

Bir numune üzerinde aynı şartlarda yapılan analizlerden aynı sonuçlar elde edilemez. Sonuçlar belli bir aralığa dağılmış halde bulunurlar. Bunun nedeni, rastgele hatalardır.

Rastgele hatalar, kontrol altına alınamayan veya birbirine bağımlı olmayan hatalardır. Böyle bir hata,

$$Er = \bar{x} - \mu \tag{1.2}$$

şeklinde gösterilir. Burada ortalama değer, μ ise istatistik ortalamadır. Rastgele hatalar iki yönlüdür. Bulunan bir sonuç, doğru değerin altında veya üstünde olabilir. Eğer bulunan sonuç doğru değerin altında ise, rastgele hata eksi, üstünde ise rastgele hata artıdır. İstatistik bilimine göre, sonsuz sayıda analiz sonucu alınırsa, bunların ortalaması doğru değerdir. Çünkü analizlerde yapılan artı ve eksi rastgele hatalar birbirini yok eder. Bunun için de, seçilen metotla yapılan analizlerde sistematik bir hatanın olmaması gerekir. Ancak pratikte yapılacak analiz sayısı hiçbir

zaman istatistikçinin istediği gibi sonsuz sayıda olamaz ve genelde en fazla 20 – 30 ile sınırlı kalır. Bu nedenle 20 analizin ortalamasına da doğru değer denir ve μ ile veya X_t ile gösterilir.

Sistematik hatalar, rastgele hataların aksine, aynı şartlarda yapılan deneylerde, aynı büyüklükte tekrarlanan hatalardır. Bunlar tek yönlüdür. Bir analizde ya daima artı olarak veya daima eksi olarak tekrarlanırlar. Bu şekilde bir yöntemin artı veya eksi hatasına o yöntemin sistematik hatası veya biası denir. Sistematik hatalar veya bias, rastgele hataların aksine kontrol edilebilen veya düzeltilebilen hatalardır. Bir yöntemin sistematik bir hatasının olup olmadığının tespiti ise şöyle yapılabilir. X maddesinin tayini için yeni bir metot geliştirildiğini kabul edelim. İçinde X maddesi bulunan bir standart numune alınır ve X maddesinin bu yeni metot ile örneğin 20 defa tayini yapılır. Elde edilen sonuçların ortalaması alınır (μ_B). Daha sonra, aynı standart numune bu defa X maddesinin tayininde biasın olup olmadığı bilinen başka bir metot ile yine 20 defa tayin edilir ve ortalaması alınır (μ_A). Buna göre yeni metodun biası;

$$\text{Bias} = \mu_B - \mu_A$$

olur. Bu değer sıfırdan farklıysa metodun belirli bir ihtimalle biası vardır denir.

Bias veya sistematik hataların meydana gelmesinin başlıca üç kaynağı vardır.

- Yöntem Hataları
- Analizci Hataları
- Enstrümantal Hatalardır.

Yöntem Hataları, yöntemde kullanılan maddelerden veya oluşan reaksiyonlardan ileri gelir. Kullanılan kimyasal maddelerin istenen saflıkta olmamaları, reaksiyonların istenilen oranda bir tarafa cereyan etmemeleri başlıca yöntem hatalarıdır. Daha detaylı yöntem hataları aşağıdaki gibi sıralanabilir.

- Maddelerin kimyasal safsızlıkları
- Kimyasal reaksiyonların yavaş meydana gelmesiyle dengenin kurulamaması
- Analizi yapılan maddenin adsorplanması veya adsorplaması
- Ayıraçların dayanıksız olması
- Buharlaşma sonucu meydana gelen kayıplar
- İstenmeyen kimyasal etkileşimler

Yöntem hatalarının kaynağını bulmak genelde, analiz veya enstrümantal hataları bulmaktan daha zordur. Bu nedenle, böyle hataları bulmak için en iyi yol,

analizi yapılacak numuneyle fiziksel ve kimyasal bakımdan aynı olan veya yakın bir standardın analiz edilebilmesidir. Standart maddelerin bileşenlerinden bazıları büyük bir doğrulukla bilinebilmektedir.

Analizci Hatalarının büyük bir yüzdesi, analizcinin çalışması esnasında yaptığı tahminlerden ileri gelmektedir. Bunlar aşağıdaki gibi sınıflandırılabilir.

- İki çizgi arasında bulunan bir cihazın göstergesinin yeri
- Bir bürette, pipette veya volumetrik bir kapta bulunan sıvının seviyesi
- Eşdeğerlik noktasında rengin şiddeti veya göstergenin yeri
- Rakamları okuma

Bazı analizciler göstergenin yerini devamlı olarak yüksek, bazıları da devamlı olarak düşük okuma eğilimi gösterirler. Bazılarının gözleri hassastır, bazılarının değildir. Hatta bazıları renk körüdür. Üniversal denecek derecede yaygın olan bir personel hatası da tekrar deneyleri sonuçlarını, bir önceki deney sonucuna daha yakın olacak şekilde okumaktır. Bütün bunlara personel hatası denir. Bias analizcide görülen bir tür hastalıktır. Analizcinin böyle hatalardan kurtulması ve analiz esnasında daima tarafsız olabilmesi için, okuduklarını en az iki defa okuması gerekir. Benzer şekilde rakamları yazarken de en az iki defa kontrol etmesi uygun olur. Analitik işlemlerde otomatik sistem ve bilgisayar kullanılması yukarıda sayılan analizci hatalarını büyük ölçüde önlemektedir.

Enstrümantal Hataların başlıcaları elektronik devrelerdeki kayma, sıcaklık değişmelerinin dedektöre etkisi, kullanılan bataryaların zamanla potansiyellerinde görülen düşmeler, güç yollarında alternatif akımın meydana getirdiği indüksiyon akımları, kullanılan kütlelerde, volumetrik kaplarda ve diğer ölçme cihazlarındaki kalibrasyon hatasıdır. Enstrümantal hatalar ve biaslar, genel olarak standartlar kullanılarak tespit ve bertaraf edilebilirler. Bu nedenle cihazların periyodik olarak kalibre edilmeleri gerekir. Çünkü cihazların cevapları, parçaların eskimeleri, korozyona uğramaları ve kötü kullanılmaları sonucu büyük ölçüde değişir.

1.4 Kesinlik ve Doğruluk

Kesinlik ve doğruluk birbirinden farklı kantitatif terimlerdir. Bir analizde elde edilen sonuçların birbirine yakınlığına veya ortalama değerle herhangi bir değer arasındaki farka **kesinlik** denir. **Doğruluk**sa, analitik yoldan bulunabilecek en güvenilir değerle doğru değer veya doğru kabul edilen değer arasındaki farktır. Biraz farklı bir şekilde doğruluk, deneyle bulunan bir sonucu doğru değere yakınlık derecesi olarak da tanımlanabilir. Bir analiz sonucunun doğruluk derecesinin hesaplanması yapılacak işlerin en önemlisidir. Bunun hesaplanmasıyla genelde ke-

sinliğin hesaplanmasına bağlıdır. Pratikte 20 veya daha fazla sonuçtan hesaplanan ortalama, doğru değer olarak kabul edilir. Tek analiz yapıldığı zaman kesinlikten söz edilemediği halde, doğruluktan söz edilebilir. Çünkü her analizde bir hata payı vardır. İki analiz yapıldığı zaman hem doğruluktan hem de kesinlikten söz edilebilir. İkinci analiz, birinci analizi kontrol amacıyla yapılır. Doğruluk, yukarıda da belirtildiği gibi bulunan bir değerin, doğru değere yakınlığının bir ölçüsüdür. Doğru değer istatistikçinin tarif ettiği teorik bir değerdir. İstatistikçiye göre doğru değer, sonsuz miktarda numune üzerinde, sonsuz zamanda, sonsuz sayıda tekrar deneyinde alınan sonuçların ortalamasıdır. Böyle bir ortalamaya popülasyon ortalaması veya sınır ortalaması denir ve μ simgesi ile gösterilir. Ancak pratikte bu kadar çok analiz yapmak mümkün olmaz. Bu nedenle, doğru değeri bulmak için yapılacak olan analiz sayısı genel olarak 20 – 30 ile sınırlıdır. Pratikte en az 20 tekrar deneyinden alınan sonuçların ortalamasına da doğru değer veya doğru kabul edilen değer denir ve Xt simgesiyle gösterilir.

$$Xt = \mu$$

Doğruluk analitin bilinen miktarlarını içeren numunelerin tekrarlanan analizleri ile tayin edilir ve aynı zamanda derişim başına minimum olarak beş tayin kullanılarak ölçülmelidir. Beklenen derişimlerin aralığında minimum üç derişim önerilir. Ortalama değer LLOQ (miktarın en düşük limiti) de beklenen asıl değerin %15 içerisinde olmalıdır ve bu değerde %20'den daha çok sapma göstermemelidir. Gerçek değerden ortalamanın sapması doğruluğun ölçülmesine hizmet eder.

Bir yöntemin doğruluğunu belirlemek için aşağıda verilen birkaç yol kullanılır.

- Saf ve karakterize edilmiş referans standart madde işlemde kullanılır ve geri kazanım hesaplanır
- Deneyde elde edilen sonuçlar valide edilmiş ikinci bir yöntemden elde edilen sonuçlar ile karşılaştırılır
- Standardın değişen derişimlerini ortama vermek ve ortamda hiç analit olmadan "placebo" denilen matriksi hazırlayıp analizi gerçekleştirmek

1.5 Sistematik Hatalar Nasıl Azaltılabilir?

Sistematik hatalar uygulanacak çeşitli metotlarla ortadan kaldırılabilirler. Bunlardan önemlileri aşağıda verilmiştir.

Şahit (Kör) Deney Metodu: Bu metot özellikle sabit hataları yok etmek için kullanılır. Bu amaçla, tayini yapılacak olan madde ortamda olmadığı halde, ortamdaymış gibi hareket edilir ve analiz gerçekleştirilir. Elde edilen analiz sonucu gerçek tayinlerde düzeltme faktörü olarak kullanılır. Bu metodu bir örnekle