

Enformetrimin İstatistiksel Temeli: Vardamlı (Inferential) İstatistik Teknikler

Özlem Gökkurt*

Enformetri, üretilmiş ve biriktirilmiş bilginin taşıyıcısı olan bütün biçimleri ile bilgiyi talep eden kullanıcılar arasında oluşan ilişkilerin sayısal çözümlemesi olarak tanımlanabilir. Bu bağlamda enformetriyi kütüphanecilik, dokümantasyon ve enformasyon biliminde matematik ve istatistiğin uyarlanmış tekniklerinin kullanıldığı bilimsel alan olarak görebiliriz. İstatistik, bir bilim dalı olarak enformetrimin iki temel yapısından birini oluşturur. Enformetrik analiz ve yasaların anlaşılabilmesi ve gündelik yaşama uygulanabilmesi için istatistiksel kavram ve tekniklere genel bir giriş yapmak zorunludur.

Genel Bir Yaklaşım

Makalemizde temelde vardamlı (inferential) istatistik olmak üzere kimi istatistik tekniklerinin, kütüphanecilik ve enformasyon biliminde araştırmalara ilişkin verilerin analizinde kullanılmasını göstermeye çalışacağız. Ele aldığımız istatistik konuları, orta düzeyde bir istatistik ya da istatistiğe giriş niteliği taşıyan eserlerde teknik düzeyde bulunabilir. Türkçe istatistik literatüründe istatistiğe giriş türü eserlerde vardamlı (inferential), betimsel (descriptive) istatistik ve olasılık (probability) teorisine ilişkin yöntem ve teknikler bulunmaktadır. Kütüphanecilik ve enformasyon biliminde enformetrimin istatistik temelini oluşturan kimi teorik dağılım ve tekniklerden sıkça kullanılan "uyumun iyiliği testi (goodness of fitness)", "ANOVA testi (Analysis of variance)" ve regresyon benzeri teknikler kullanılmaktadır. Bunlara ek olarak "boyutsal indirgeme teknikleri (dimensional reduction techniques)" ne özellikle enformetrik yasaların kurulmasında çok değişkenli kütüphanecilik ve enformasyon teorilerine uygun modellerin analizinde ileri düzeyde yer verilmektedir. "Faktör analizi", öbikleme analizi (cluster analysis)" ve "çok boyutlu ölçüm (multidimensional scaling)", boyutsal indirgeme tekniklerinin sık rastlanan örnekleridir. Daha çok bilgi erişim (information retrieval) ve bilgisayar temelli indeksleme işlemlerinde bu ileri veri analiz teknikleri kullanılmaktadır. Yazımızın konusu yukarıda saydığımız istatistik tekniklerinin özel kütüphane ve enformasyon işlemlerine uygulanması değildir. Ülkemizde betimsel istatistik tekniklerini yani aritmetik ortalama ve yüzde işlemlerini aşabilen vardamlı ve ileri düzey istatistik tekniklerinden yararlanan kütüphanecilik ve enformasyon bilimleri araştırmalarına rastlayamamaktayız. İstatistik tekniklerinin kısa tanıtımının bize bu yolda Türkiye için gecikmiş bir alana girme olanağını sağlayacağını umuyoruz.

* DTCF Kütüphanecilik Bölümü Araştırma Görevlisi

İstatistikte Temel Ayrımlar

İstatistik, işlenmiş veri yığınının temel niteliklerini temsil etmek, özetlemek ve diğer ileri düzey istatistik tekniklerle kullanılmak üzere veri yığınını tanımlayıcı sayısal değerler hazırlayan yöntem, teknik ve işlemleri içerir.

Bu yaklaşıma göre; istatistik, genelde iki temel bölüme ayrılmıştır: Betimsel (Descriptive) ve vardamalı (inferential) istatistik.

Vardamalı istatistik, bir yığından alınmış örneklemden yola çıkarak yığının genel karakteristikleri, olası gelişmeleri ve değişkenler arasındaki ilişkileri test eder. Örnek olarak enformasyon biliminde verilmiş bir alanda her bir makalenin tek tek analizi ile yığın olarak düşünülebilecek olan verilmiş alanın tüm makalelerinin analizini yapmak kütüphaneci için imkansızdır. Bu durumda vardamalı istatistik tekniklerine uygun olarak verilmiş alandan örneklem yapmak suretiyle belirli sayıda makale alıp yine diğer vardamalı istatistik teknikleri kullanılarak bütün yığına ilişkin karakteristikler, olası gelişmeler ve eğilimler hesaplanabilir.

İstatistik diğer bir açıdan tek değişkenli, iki değişkenli ve çok değişkenli istatistik olarak üç temel parçaya ayrılabilir. Tek değişkenli istatistik, tek bir değişkenin karakteristiklerini betimlemeye yöneliktir. Genellikle betimleyici istatistik, tek bir değişken üzerinde çalışır. İki değişkenli ve çok değişkenli istatistik, iki veya ikiden daha fazla değişken arasındaki ilişkileri tanımlamaya yönelik teknikleri veya böyle bir ilişkinin varlığı, yokluğu konusunda geliştirilmiş olan hipotezleri test etmeye yönelik istatistiksel teknikleri içerir. Tek değişkenli betimsel istatistik, merkezi eğilim ölçüleri ve dağılım ölçüleri üzerinde çalıştığından ileri analiz tekniklerine ve değişkenler arasındaki ilişkilere dayalı enformetrik modellerde kesinlikle yardımcı olamaz. Kısacası, enformetrik modeller ve o modellerden türetilmiş olan enformetrik yasalar, vardamalı istatistik tekniklerinin bir bütün olarak kullanılmasını zorunlu kılmaktadır. Tüm vardamalı istatistikler, iki değişkenli ya da çok değişkenli özellikler üzerinde çalışmaktadır.

Vardamalı istatistik, iki temel değişken türü üzerinde durur. Birbirlerine çok yakın olan istatistik tekniklerinin hangisinin, hangi koşullarda nasıl uygulanması gerektiği, yukarıda söz ettiğimiz iki değişken türünün açıklıkla belirlenmesine bağlıdır. Birinci değişken türü sürekli (continuous) değişken türüdür. İkinci değişken türü kesikli (discrete) değişken türüdür.

Olasılık dağılımlarında binominal, poisson, normal dağılımlardan hangisinin tercih edilmesi gerektiği, yukarıda söz ettiğimiz veri türleri (yapıları)ne bağlıdır. Bu veri türlerinin kullanımında üç olasılık olabilir:

1. Değişkenlerden birisi kesikli, diğeri süreklidir.
2. Her iki değişken de süreklidir.
3. Her iki değişken de kesiklidir.

Kütüphanecilik ve enformasyon alanındaki verilerin ölçüm ve sayım durumlarında kesikli ve sürekli değişkenlerle hesaplamalar yapabileceğimizi düşünecek olursak, vardamalı istatistiğin ve teorik dağılımların ilgili tekniklerini kullanmamız gerekecektir.

Bir başka açıdan değişken türlerini incelediğimizde, değişkenlerden birisi bağımsız ya da belirleyici değişken iken diğeri bağımlı ya da belirlenen değişken olarak sınıflandırılabilir. ANOVA benzeri teknikler (student t testleri ve bazı non-parametrik testler) ölçümlü belirtilen sürekli bir değişken yönünden ikiden çok grup arasında fark olup olmadığını test etmek için

kullanılır. Örneğin, bir kataloglama bölümünde çalışan üç katalogçunun farklı konulardaki kitapları sınıflandırma işlemindeki verimlilikleri ortak bir bazda ANOVA teknikleriyle test edilebilir.

Regresyon ve ilgili teknikler, her iki değişken de süreklilik gösteren değişken ölçümleri üzerinde uygulanır. Bu tekniğin amacı, değişkenler arasındaki ilişkinin anlamlılığına ve bağımsız değişkenin alacağı her bir farklı değer için bağımlı değişkenin alacağı değerleri hesaplamakta kullanılır. Kütüphanecilik alanından örnek vermek gerekirse, bir kütüphanede ekonomi alanında ödünç verilen süreli yayınlarla aynı konudaki ödünç verilen kitaplar arasında bir ilişkinin olup olmadığı; eğer anlamlı ise kaç kitapla kaç süreli yayın ödünç verildiğinde ne kadar kitabın ödünç verilebileceğini istatistiksel olarak tahmin etmek mümkündür.

Regresyon teknikleri ilişkilerin varlığının analiz edilmesinde, ANOVA teknikleri ise karşılaştırmaların yapılmasında kullanılır.

Özet olarak ANOVA, regresyon teknikleri iki değişkenli durumlar için temel istatistiksel araçlar olup, vardamlı istatistiğin en önemli parçalarıdır. Üç ve daha fazla değişken içeren durumlarda yani çok değişkenli ilişkiler yapısında çok değişkenli vardamlı istatistik teknikleri kullanılabilir.

Kütüphanecilik ve enformasyon biliminde veri analizi için istatistik tekniklerinin kullanımı uluslararası literatürde eski olmasına rağmen, Türkiye’de oldukça yenidir. Ampirik araştırma ve istatistiksel değerlendirme, toplumbilimleri, doğal bilimler ile teknoloji alanında kullanıldığı gibi alanımızda da kullanılmak durumundadır.

Enformasyon biliminde ölçülebilir nesnelere, üç temel sınıflandırılmasından söz edilebilir:

1. Üretilmiş/biriktirilmiş bilgiler
2. Sistemler
3. Kullanıcılar

Üretilmiş ve biriktirilmiş bilgilerin var oluş biçimleri belgeler ve belge benzeri kayıtlardır. (Kitap, dergi, mikroformlar, CD Romlar vb.). Enformasyon bilimi, bilginin özünüyle ilgiliyken kütüphanecilik var oluş biçimiyle ilgilenmektedir. Her iki tür de birbirlerine bağlı olarak ölçülebilir nesnelere çalışırlar. Enformasyon sistemleri, ölçülebilir olanlardır. İstatistiksel teknikler bu alanda sistemi analiz etmek için kullanılmaktadır. Nihayet kullanıcı analizi, sistemi kullanan kullanıcının kimi ölçütler baz alınarak sistem karşısındaki karakteristik özellikleri ve davranış biçimlerini belirlemeye yönelik istatistiksel çalışmalardır.

Kısaca, sayılabilir ve ölçülebilir konular üzerinde çalışmalar yapan bilim dalımız istatistik tekniklerini toplumbilimleri alanlarına göre çok daha kolay ve kesin biçimde uygulayabilmektedir.

Enformetri, enformasyon biliminin tüm alanlarında istatistik tekniklerinin tek başına yeterli olamayacağını söylemektedir. Bazı alanlarda matematiksel formülasyonların ve ileri analiz biçimlerinin daha uygun olduğu görülmektedir. Bir enformasyon Sisteminin algoritmasının biçimsel özelliklerine yönelik hesaplamalar ya da sistem davranışının teorik modelinin geliştirilmesi, sayısal bilgiler üzerinde çalışılacak alanlar olmakla birlikte matematiksel çözümlenmeyi gerektirirler, fakat istatistiksel değildirler.

Sonuç olarak enformetride istatistik teknikler, betimsel olandan vardamlı olana, basitten

karmaşığa bir bütün olarak kullanılmaktadır. Tek tek kütüphane ve enformasyon merkezlerinin davranışlarından ölçüm ve sayımlar yapmak suretiyle genel enformasyon yasalarına ve enfometrik modellere ulaşmak mümkündür.

Kaynaklar

Bennion, Bruce C.-S. Karschamroon. "Multivariate regression models for estimating journal usefulness in physics." *Journal of Documentation*, 40:3(1984)217-227.

Egghe, L.-R. Rousseau. *Introduction to informetrics: Quantitative methods in library, documentation and information science* (Amsterdam:Elsevier,1990)

Gökkurt, Özlem. "Enfometri (Informetrics): Bir tanıtım denemesi" *Kütüphane Enformasyon Arşiv Alanında Yeni Teknolojiler ve TÜRCMARC Sempozyumu Bildirileri 14 Ekim 1991* (Yay. Haz. Hasan S. Keseroğlu, İstanbul: Türk Kütüphaneciler Derneği İstanbul Şubesi,1991).

Gökkurt, Özlem. *Özel kütüphanelerde kullanıcı analizi sorunu*. Ankara Üniversitesi Sosyal Bilimler Enstitüsü Kütüphanecilik Ana Bilim Dalı Yüksek Lisans Tezi (Ankara, 1991).

Kinnucan, Mark T.-Michael J. Nelson-B.L.Allen. "Statistical methods in information science research." *Annual Review of Information Science and Technology* (ARIST), 22 (1987) 147-178.

Pontigo, J. F.W.Lancaster. "Qualitative aspects of the Bradford distribution." *Scientometrics*, 9:1-2(1986).