Sosyal Bilimler Dergisi 2008, (1), 15-18

KOŞULLU OLASILIĞIN DİLİ

Çeviren: Yard. Doç. Dr. L. Aylin AKTÜKÜN

İ. Ü. İktisat Fakültesi, Ekonometri Bölümü, Yöneylem Anabilim Dalı

ABSTRACT

Statistical terms are accurate and powerful but can sometimes lead to misleading impressions among beginning students. Discrepancies between the popular and statistical meanings of "conditional" are discussed, and suggestions are made for the use of different vocabulary when teaching beginners in applied introductory courses

Key Words: Statistics education, Statistical language

ÖZET

İstatistik terimler kesin ve güçlüdürler, ancak bazen yeni başlayan öğrenciler için yanıltıcı izlenimlere neden olabilirler. Bu çalışmada "Koşullu" kelimesinin popüler ve istatistiksel anlamları arasındaki farklılıklar tartışıldı ve yeni başlayanlara uygulamalı giriş derslerinde kullanılmak üzere bu iki farklı kullanım için öneriler sunuldu.

Anahtar Kelimeler: İstatistik Eğitimi, İstatistiksel Dil.

1. GİRİŞ

Alice "'ihtişam' ile neyi kastettiğini anlamadım" dedi.

Humpty Dumpty kibirli bir şekilde güldü. "tabii ki anlamazsın- ta ki ben sana söyleyene kadar. Kastettiğim, seni yere serebilecek güzel bir savım söz konusu!".

Alice "ama 'ihtişam' beni yere serecek güzel bir sav anlamına gelmez ki" diyerek itiraz etti.

"Bir sözcük kullandığım zaman, hangi anlama gelmesini seçtiysem o anlama gelir- ne daha fazla ne daha az" dedi Humpty Dumpty küçümseyen bir tonla.

Lewis Carroll, Through the Looking Glass, The Dial Press, Toronto, 1931. p. 238.

Humpty Dumpty gibi, istatistikçiler de genellikle sözcükleri hangi anlama gelmek üzere seçtilerse o anlamda kullanırlar. Yani bir istatistikçiye göre "önemli" nin anlamı artık "mühim" değil, "sıfır hipotezi altında muhtemel olmayan" anlamına gelecektir. Alice gibi öğrenciler de kavramları yeni bir dilde kullanarak anlayabilirler.

Bunun gibi özelleştirilmiş bir istatistiksel kelime dağarcığı uzmanlar için güçlü bir araç olabilir, ancak uygulamalı istatistik derslerinde bildik/aşina olunan sözcükleri kullanmak, yeni başlayanların yeni fikirleri anlamalarında daha fazla işe yarayabilir. Eğer öğretmenler "B verilmişken A" gibi ahşılmadık teknik bir ibareden kaçımp, onun yerine küme teorisi bağlamında yorumlama şansı veren "A, B'ye dahil" şeklinde betimleme yaparlarsa yanıltıcı olabilen koşullu olay kavramını daha kolay izah edilebilirler. Küme teorisi yorumlarını daha fazla teşvik edici hale getirmek için öğretmenler koşullu olasılıkları "altküme olasılıkları" olarak tanımlayabilirler. Anlaması kolay olan küme teorisi dilini sunmak, öğrencilerin önemli kavramlara odaklanmasını ve koşullu olasılıkla ilgili çok yaygın hatalardan kaçınmalarını sağlayabilir.

2. KOŞULLU OLASILIKLARLA İLGİLİ YAYGIN HATALAR

Öğrenciler tarafından yaygın olarak yapılan hatalardan biri P(A|B) ile P(B|A)'yı karıştırmaktır, şöyle ki, örneğin, bir pozitif tanı (bir testin pozitif kestirim gücü) söz konusuyken hastalığın gerçekte olmasına ilişkin koşullu olasılığı, hastalık gerçekte varken testin pozitif tanı koyması koşullu olasılığı (testin duyarlılığı) ile aynı gibi düşünülmektedir. Diğer bir yaygın hata da P(A|B) ve P(A) arasındaki farkın anlaşılmaması hatasıdır; test sonucu pozitifken hastalığa sahip olma olasılığının, hastalığa sahip olma olasılığı (bu hastalığı taşıyanların oranı) ile aynı olduğunun düşünülmesidir.

Bu hatalar yine aynı temel problemden kaynaklanmaktadır; örneklem uzayının ya da frekans hesabının paydası değiştiğinde. A|B, B|A ve A olaylarının tümü aynı olay olarak düşünülmekte ve öğrenciler bunları aynı tekniklerle betimlemektedirler. Öyleyse etkin bir öğretim için olayların farklılığını yaratan özelliklere dikkat çekmelidir. "B verilmişken A" ibaresi yeni bir olaylar kategorisinin varlığına dikkat çekmede yeterli olmayabilir, bunun nedeni belki de bu ibarenin istatistiğin dışında nadiren kullanılmasıdır. Koşullu olayları "A, B'ye dahil" ya da "A, B'nin içinde" şeklinde tanımlamak öğrencileri bu olayları birbirlerinin alt kümesi olarak algılamalarına ve böylece yeni bir kategorinin varlığını altının çizilmesine yardım edebilir. Böylelikle bu dil olasılık kitaplarında koşullu olasılığı öğretmek için sıkça kullanılan "indirgenmiş örneklem uzayı" kavramını daha iyi anlamlandırır (Ross 2002).

Koşullu olasılıkları "A, B'nin içinde" şeklinde betimlemek özellikle Venn diyagramları ve kontenjans tablolarına başvurulduğunda faydalıdır. Her ikisi de olay uzayları ve olasılıkları görselleştirmede güçlü araçlardır. Bunun nedeni muhtemelen öğrencilerin bunları olasılıktan ziyade frekans oranları olarak kavramsallaştırınasına ve dolayısıyla daha kesin yargılara ulaşmasını sağlamalarıdır (Cosmides and Tooby 1996; Hoffrage, Lindsey, Hertwig and Gigerenzer 2000).

3. KOŞULLU OLAYLARI ARDIŞIK OLAYLAR OLARAK DÜŞÜNMEKTEN KAYNAKLANAN PROBLEMLER

Bazı öğrenciler için "koşullu" sözcüğünün kendisi yanıltıcıdır, çünkü yaygın söylemde genellikle bir olaydan önce başka bir olayın gerçekleşme zorunluluğu gibi bir olguya dikkat çeker. Örneğin, bir üniversiteye "koşullu kabul edilen" bir öğrenci, kampüste boy göstermeden önce liseyi bitirmek zorundadır. Öğrenciler "koşullu" sözcüğünü istatistik bir içerikte (kendi istatistiksel anlamında) ilk kez duyduklarında, koşullu olayları bu örnekteki gibi ardışık olaylar olarak düşünebilirler.

Ancak istatistikteki koşullu olaylar ardışık olaylar gibi kavranıldığında bazen kafa karıştırıcı olabilirler. Bir bioistatistik kitabından alınan aşağıdaki örneği düşünelim (Pagauo and Gauvreau 2000): ABD'de belirli bir yıla ilişkin hayat tabloları verisi erkek bebek doğma olasılığını 0.512 olarak göstermektedir. Bu topluluktan iki hamile kadın rastlantısal olarak seçilsin. En az birinin erkek çocuk doğurduğunu biliyorsak, her ikisinin de erkek çocuğu doğurmuş olma olasılığı nedir?

A olayı, en az birinin erkek çocuk olma olayı, $b1$ ve $b2$, 1. ve 2. çocukların erkek olması, $g1$ ve $g2$ ise 1. ve 2. çocukların kız olması olayları ise, problem koşullu olasılık tanımıyla aşağıdaki gibi çözülebilir:

$$P(b1 \cap b2 | A) = \frac{P((b1 \cap b2) \cap A)}{P(A)}$$

$$= \frac{P(b1 \cap b2)}{\{P(b1 \cap b2) + P(b1 \cap g2) + P(g1 \cap b2)\}}$$

$$= \frac{0.262}{\{0.262 + 0.250 + 0.250\}} = 0.344$$

Ne yazık ki eğer öğrencileri olayların ardışıklığına odaklanmaya teşvik edersek, bir çocuğun cinsiyeti sihirli bir biçimde diğer çocuğun cinsiyetini etkiliyor gibi görünür. Hatta formülü doğru uygulayan öğrenciler bile rastlantısal olarak seçilen doğumların bağımsız oldukları için birbirlerini etkileyemeyeceklerini bildiklerinden çözümlerinin yanlış olduğunu düşünme eğiliminde olabilirler. (Gerçekten $b1$ ve $b2$ bağımsızdır, ancak $(b1 \cap b2)$ ve A değildir.)

Küme teorisi dili bir doğumdaki cinsiyetin bilinmesinin, örneklem uzayının indirgenmesi olduğuna dikkati çekerek, olayların zaman içindeki ardışıklığının önemini azaltır. Bunun gibi problemler için "B verilmişken A" koşullu olasılığı "A'nın B'ye dahil olma" olasılığı olarak ifade edilebilir. Problem şöyle ifade edilebilir: "en az bir çocuğun erkek olduğu alt kümeler içinde her iki çocuğun da erkek olma olasılığı nedir?"

4. ALT KÜME OLASILIKLARI VE P-DEĞERLERİ

P-değerleri anlatılırken koşullu olasılıkları alt küme olasılıkları olarak betimlemek oldukça faydalıdır. Ders kitapları genellikle P-değerini sanki sıfır hipotezi bir şemsiyeymiş gibi "sıfır hipotezi altındaki olayın olasılığı" olarak betimlerler. Ancak öğrenciler "sıfır hipotezi altındaki" gibi garip bir söylemi göz ardı edip, P-değerine basitçe "olayın şans eseri ortaya çıkma olasılığı" olarak düşünme temayülündedirler.

Şayet öğrenciler, P-değerinin tümleyenini sıfır hipotezinin yanlış olma olasılığı ya da alternatif hipotezin doğru olma olasılığı olarak düşünürlerse, bu bakış açısı p-değerinin problemli/yanlış kavranılmasına yol açabilir. (Rossman and Short 1995). Örneğin bazı öğrenciler şöyle düşünebilirler: "iki ortalama arasındaki farka ilişkin P-değeri = 0.01 ise, bu farkın şanstan kaynaklanmama olasılığı %99'dur." Öğrenciler P-değerini sıfır hipotezinin betimlediği bir dünyadaki olayın olasılığı olarak öğrendikleri zaman, tümleyeninin alternatif hipotezin olasılığı olduğu yanılgısına genellikle düşmezler.

5. STANDART KELİME DAĞARCIĞINA GEÇİŞ

Sınıfta koşullu olasılık ilk kez anlatıldığında P(A|B)'nin "A'nın B'ye dahil olma olasılığı", yani bir alt küme olasılığı olarak çevrilmesi faydalıdır.

Bu söylem, koşullu olasılıkta örneklem uzayına ve onun koşullu olasılıkla nasıl indirgendiğine dikkati çeker. Ancak öğrencilere, bu yapıya alışıldıktan sonra "B verilmişken A'nın koşullu olasılığı" gibi standart söylemler sunulabilir, böylelikle öğrenciler ders kitaplarındaki ve istatistik literatüründeki dile aşina hale gelebilirler.

6. SONUÇ

Öğrenme üzerine oldukça etkin olan bir rapora göre "Öğretmenler, öğrencilerinin daha önceki kavrayışlarına ilişkin bir liste oluşturup, üzerinde çalışmalıdırlar" (Bransford, Brown and Cocking 2000). Öğretmen daha önceki bu kavrayışları bir kere anlarsa, bunlarla sınıfta sunulacak yeni fikirler arasındaki bağlantıları kurarak öğrenmeyi kolaylaştırabilir (Knowles 1988; Lovett and Greenhouse 2000). Örneğin, bir öğretmen tıp ya da diğer sağlıkla ilgili bölüm öğrencilerine Bayes Teoremini, öğrencilerden yanlış pozitif ve yanlış negatif test sonuçlarına ilişkin kişisel deneyimlerini hatırlamalarını isteyip, bunları koşullu olasılık hesabı ile ilişkilendirerek öğretebilir.

Bir öğretmen, öğrencilerin kelime dağarcıklarını daha önceki kavrayışlarının bir parçası olarak düşünmelidir. Koşullu olasılığı geleneksel istatistik terimler yerine aşina olunan bir dilde sunmak –özellikle bu dil, küme teorisi bağlamında koşullu olayların görsel bir temsilini sağlıyorsa, varolan bilgileri ilişkilendirmenin bir yolu olabilir. Küme teorisi söylemi, geleneksel istatistik kitaplarla ya da koşullu olasılığı basitleştirmek amacıyla geliştirilmiş diğer pek çok zarif ve faydalı öğretme yöntemleriyle birlikte kullanılabilir (Warner, Pendergraft and Webb 1998; Hoffrage, et al. 2000).

İstatistiğe giriş dersi alan öğrenciler kendilerini sıkça istatistiğin kafa karıştırıcı harikalar diyarındaki Alice'leri gibi hissederler. İstatistik öğretmenleri aşina olunan bir dil kullanarak yeni kavramların gerçek dünya ile ilişkilerini göstererek, bu karışıklığı önleyebilirler.

Teşekkür

Yazar, Robert Wood Johnson/National Library of Medicine (3T15-LM007079-14S1 in public health informatics) tarafından bir doktora öncesi bursuyla desteklenmiştir.

KAYNAKÇA

Bransford, J.D., Brown, A.L., and Cocking, R.R. (eds.) (2000), **How People Learn:Brain, Mind, Experience,** *and School,* Washington, D.C.: National Academy Press.

Cosmides, L. and Tooby, J. (1996), "Are Humans Good Intuitive Statisticians after All? Rethinking Some Conclusions from the Literature on Judgment under Uncertainty", **Cognition**, 58, 1-73.

Hoffrage, U., Lindsey, S., Hertwig, R., and Gigerenzer, G. (2000), "**Communicating Statistical Information", Science**, 290, 2261-2262.

Knowles, M.S. (1988), **The Modern Practice of Adult Education: From Pedagogy to Andragogy**, Englewood Cliffs, N.J.: Cambridge Adult Education.

Lovett, M.C. and Greenhouse, J.B. (2000), "Applying Cognitive Theory to Statistics Instruction", **The American Statistician**, 54, 196-217.

Pagano, M. and Gauvreau, K. (2000), **Principles of Biostatistics** (2nd ed.), Pacific Grove (CA): Duxbury Thompson Learning.

Ross, S. (2002), **A First Course in Probability** (6th ed.), Englewood Cliffs, NJ: Prentice Hall.

Rossman, A. J. and Short, T.H. (1995), "Conditional Probability and Education Reform: Are They Compatible?" **Journal of Statistics Education** [Online], 3(2).

www.amstat.org/publications/jse/v3n2/rossman.html

Warner, B.A., Pendergraft, D., and Webb, T. (1998), "That Was Venn, This Is Now", **Journal of Statistics Education** [Online], 6(1).

www.amstat.org/publications/jse/v6n1/warner.html