

Ölçmelerde Hata, Yanlılık ve Diferansiyel Madde Fonksiyonu

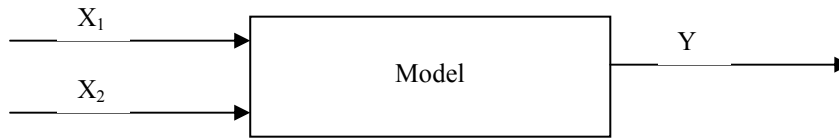
Hata Kuramı

Hata kavramının iki tanımı yapılmaktadır. Bunlardan ilki; gerçek değer ile ölçülen değer arasındaki farklılık, diğeri ise bir deneydeki belirsizlik olarak adlandırılan ve standart sapma gibi nicel değerler ile ifade edilebilen durumdur (Beers, 1957: 3).

Bir soru ve bir yanıt daha açıklayıcı olacaktır.

Soru: $1+1$ neden 2 eder.

Yanıt: Sistem ya da model aşağıdaki gibi tanımlandığından dolayı.

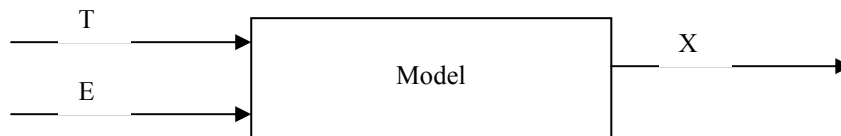


Şekil 1: Deterministik model

Bilindiği gibi, matematiksel (deterministik) modeller analitik (0 hatalı) çözüm üretirler. Bu nedenle genellikle belirli bir fonksiyondan türetilen “laboratuar verileri” ile çalışırlar. Gerçek yaşamdan elde edilen veriler ise hata içerirler. Bu tür veriler ile kurulan modeller ise istatistiksel (olasılıksal-probabilistic) modellerdir.

Diğeri taraftan, bir problemin olası üç çözümü vardır, bunlar sırasıyla; analitik (mutlak çözüm), iteratif (hatanın kontrol edilmesine dayalı çözüm) ve simülatif (yaklaşıklıkla çözüm) çözümlerdir. Deterministik modeller matematikte kullanılırken diğeri çözümler ise istatistiksel modellemelerde kullanılırlar ve hata terimi içerirler.

Deneyisel çalışmalarda ortaya çıkan hataların bir kısmı ölçmeciden ölçmeciye değişim gösterdiği ve kaynağı bilinmediği için rasgele hatalar (random error) olarak adlandırılırlar. Rasgele hatalar aynı zamanda modele alınmayan değişkenleri temsil etmektedir.



Şekil 2: İstatistiksel Model

X gözlemi bir sonuç ya da bir olaydır. Bu olayın/sonucun ortaya çıkmasında etken olarak rol alan değişkenler ile X sonucu açıklanmaya çalışılır. $X=Y_1+Y_2$ gibi bir modelde X olayı Y_1 ve Y_2 ile açıklanır ve bu tür analitik modellerde ortaya çıkan X değişkeninin %50'si Y_1 , geri kalan %50'si ise Y_2 tarafından açıklanır. Şekil 2'de $X=T$ olduğunda analitik bir modele ulaşılır.

Ancak modelde eksik (modele alınmamış) değişken var ise o zaman ne olur? X'nin ortaya çıkmasında etken olan ve modelde yer verilmeyen değişkenler hata (E) değişkeni olarak modele sokularak model tamamlanır. Bir diğer ifade ile E değişkeni modele alınmayan değişken(ler)i temsil eder. Bu durumda $X=T+E$ modeli istatistiksel bir model, hata (E) değişkeni ise modelin rasgele hatası olarak adlandırılır.

Özellikle eğitim bilimlerinde ölçmeye konu olan özellik tek boyutlu ise Şekil 2 ile verilen yapı geçerlidir. Eğer ölçmeye konu olan özellik iki boyutlu ise $X=T_1+T_2+E$ şeklindeki bir model üzerinde çalışmak gerekir.

Gözlenen puanlar (X) bir sonuçtur ve bu sonuç içerisindeki ölçülmek istenen gerçek puanların (T) etkisi/katkısı ise modelin ya da ölçme işleminin güvenilirliğini verir. Güvenirlik (reliability) ise regresif modelin belirtme katsayısının (determination coefficient) bir fonksiyonudur.

Rasgele hatalar, ölçme işlemini yapan kişiden kaynaklandığı için ölçme işleminin doğasında olan bir durumdur ve istatistiksel kestirimler için bu tür hatalardan yararlanılır. Bir bakıma, istatistiksel kestirimler için bu tür hatalara ihtiyaç vardır. Ancak bazı hatalar vardır ki ölçme aracından kaynaklanır ve modelde hiçbir şekilde yer verilmez. Bu tür hatalara sistematik hata adı verilir ve istatistiksel modellerin yanlış kestirilmelerine yol açar.

Baykul ve Turgut ölçmelere karışan hataları sabit, sistematik ve rasgele hata olmak üzere üç türde ele almışlardır (Baykul, 2000: 107, Turgut, 1992:28). Baykul'un ifade ettiği gibi sabit hata, ölçme aracından kaynaklanan ve her ölçme işlemine eşit miktarda karışan hatalar olarak adlandırılır. Sistematik hata ise, sabit hata ile aynı içerikte ama ölçmelere değişen miktarlarda karışmaktadır. Ancak Beers, sabit hatayı da sistematik hata olarak aynı sınıfta ele almakta ve sistematik hatayı "belirli sistematik hata" ve "belirsiz sistematik hata" olarak iki gruba ayırmaktadır (Beers, 1957: 3).

Sabit hatalar modelde temsil edilememelerine karşılık miktarı belli olduğu için ölçme sonuçlarının düzeltilmesi ile bir şekilde giderilebilmektedir. Ancak sistematik hatalar için böyle bir durum söz konusu değildir. Ölçmelere sistematik hata karışıyor ise, öncelikle bu durumun tesbiti yapılmalı daha sonra ise ölçme aracı iyileştirilmelidir.

Genel bir ifade kullanmak gerekirse, modelde yer verilmesi gerekirken modelde temsil edilmeyen değişkenler güvenilirliğin bir sorunudur. Modelde yer alması istenmediği halde bir şekilde modele karışan değişkenler ise geçerliliğin bir sorunudur.

Davranış bilimlerindeki ölçme ve değerlendirme kapsamında yapılan çalışmalarda rasgele hatanın varlığını araştırmak bir güvenilirlik çalışmasıdır, sistematik hatanın varlığını araştırmak ise yanlılık araştırmasıdır. Bununla birlikte her iki hata türü de ölçme aracının geçerliliğinin azalan bir fonksiyonudur. Ancak sistematik hata özellikle yapı geçerliliğini olumsuz yönde etkilemektedir (Zumbo, 1999:9).

Bilinen ile bilinmeyen arasındaki bağıntı aynı zamanda bilgi kuramının bir modelidir. EK'te bu konuya ilişkin bilgi verilmiştir. Şekil 2'de verilen model ile EK'te verilen model benzerlik göstermektedir. Şekil 2'de kaynaktan gelen (T) bilgi miktarı ve hedef birime (X) ulaşan özbilgi (entropi) miktarları elde edilmeye çalışılır. Şekil 2'deki T'nin sağladığı bilgi miktarı:

$$I(T) = -\log P_i \quad (1)$$

olarak ifade edilebilir. ÖÖK'nda ise bilgi miktarı,

$$I(\theta) = -E\left\{\frac{\partial^2}{\partial\theta^2} \log L\right\} \quad (2)$$

olarak verilmektedir (Hambleton ve Swaminathan, 1985:102). Bu ifade aynı zamanda Fisher bilgi fonksiyonu olarak bilinmektedir ve Fisher normal dağılıma sahip bir rastlantı değişkenine ilişkin bilgi olarak;

$$I(\mu) = \frac{1}{\sigma^2} \quad (3)$$

ve binom dağılımına ilişkin bilgi olarak ise

$$I(p) = \frac{1}{pq} \quad (4)$$

bağıntısını göstermiştir. Bu nedenle yukarıdaki ifade türleri Fisher Bilgisi (Fisher's Information) olarak kavramlandırılmışlardır (Dudewicz ve Mishra, 1988:371).

Fisher bilgisinden hareket ile; bilginin tersi belirsizliktir (uncertainty). Bir diğer ifade ile hatadır denilebilir. Bilgi arttıkça belirsizlik azalır. Bu nedenle *hata* için, yukarıdaki ifadenin tersi,:

$$\text{Var}_{\text{Hata}}(\theta) = \frac{1}{I(\theta)} \quad (5)$$

olarak belirtilebilir. Bu ifadenin karekökü, ölçmenin standart hatasını verir (Hambleton ve Swaminathan, 1985:104).

Psikolojik değişkenleri içeren ve gözlenen-gözlenemeyen arasındaki bağıntıyı açıklamaya çalışan modellerde ÖÖK Fisher'in bilgi fonksiyonu üzerine kuruludur. KTK ise Pearson'ın karşılıklı bilgi (mutually information) kuramı üzerine kuruludur. Pearson 1904 yılında ve Linfoot ise 1957 yılında bilgi miktarını;

$$I(X) = -\frac{1}{2} \log(1 - \rho^2) \quad (6)$$

şeklinde ifade etmiştir (Hamdan ve Tsokos, 1971) ve bu ifade aynı zamanda

$$\rho^2(X, T) = \frac{\sigma_T^2}{\sigma_X^2}$$

ile verilen güvenilirlik indeksinin bir kestiricisi olarak ele alınabilir.

Eşitlik 6 ile verilen ifade, Şekil 2 göz önüne alındığında; T kanalından X'e gelen bilgi miktarını belirtmektedir. Yani, belirsizlikten/hatadan arındırılmış bilgi miktarıdır ve bu ise

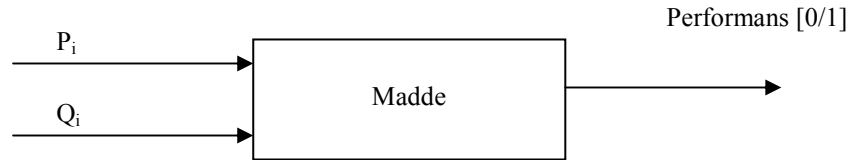
Turgut (1992:31) tarafından yapılan KTK'ında gvenirliđin tanımına karřılık gelmektedir. Bu bilgi miktarının dűřmesi, belirsizlik miktarının artmasına yol aęar.

Yukarıda bahsedilen durum, T kanalından X alıcısına gelen sinyalin/bilginin sađlıklı olması durumunda geęerlidir. Bilginin ya da bilgi miktarının bozulumu da sűz konusu olabilir. Bu durum yalnızca ۆlęmenin sistematik hatadan kaynaklanmasından oluřabilir.

Sistematik hatayı da modele katmak suretiyle KTK'na iliřkin modeli $X=T+S+E$ řeklinde tanımlamak gibi bir giriřimde ise; bilgi kuramı geređi bu gűsterim geęersiz olmaktadır. Modeldeki S'in (sistematik hata) varlıđı yalnızca $T \rightarrow X$ ve $E \rightarrow X$ arasında tařınan bilgi miktarını bozmaktadır.

Ölęmelere karıřan sistematik hatalara iliřkin bilgiler bir sonraki konuda verilecektir.

KTK ۆzerine kurulu olan madde analizleri ięinde bilgi-kuramsal yaklařım sűz konusudur. ۆrneđin, 0 ve 1 řeklinde puanlanmış maddeye iliřkin veri kűmesinde madde gűçlük indeksinin 0,5 olması istenir. Buradaki amaę bilgi miktarını enbűyűklemektir (maximized).



$P_i=1/2$ $Q_i=1-P_i=1/2$ ise ortalama bilgi miktarı;

$$H_i = \sum_{i=1}^2 P_i \log_2 P_i = (1/2) \log_2 2 + (1/2) \log_2 2 = 1 \quad (\text{bit})$$

Eřitlik 3 ve 4'te verilen bađıntılardan hareket ile $P_i=0,5$ olduđunda 1 bitlik bilgiye ulařılır ve bilgi miktarı bu noktada maksimumdur.

Sistematik Hata ve Yanlılık Kavramı

Bilgi kuramı, ilk olarak elektrik ve elektronik konularında yer alan iletiřim alanında ortaya ęıktıđı ięin bilgi, kanallar ۆzerinden geęen sinyalin tařıdıđı "bit" olarak tanımlanmakta idi. Psikolojide ise bu yaklařım organizmanın sahip olduđu hipotetik yapı ile uyarıcıya verdiđi tepki (performans) arasındaki ۆlęmelerde aranmaktadır. Buradaki bilgi bozulumu ancak

uyarısının niteliğinden kaynaklanabilir. Bilgi bozulumu, sistematik belirsizliklerden yani sistematik hatalardan (systematical error) kaynaklanır. Bu durumda Eşitlik 5 ile verilen bağıntı ile $E(X)=T$ bağıntısı ortadan kalkar ve $\delta_S=X-T-E$ ifadesine istatistiksel olarak "yan" adı verilir. Rasgele hata ise $\delta_R=X-T$ ifadesi ile açıklanabilir.

Eğitimde alanında yapılan ölçme ve değerlendirme çalışmalarında da, ölçmelere karışan sistematik hatalara "yan" (bias) adı verilmektedir. Ancak bu alanda kullanılan değişkenler psikolojik değişkenler ve ölçme aracı da *geliştirilmiş* ölçme aracı olduğu için sistematik hatanın kaynağı iyi tespit edilmelidir.

Aşağıda üç adet örnek verilmiştir.

Örnek 1: Camilli & Shepard (1994:8):

"Kronometre ile bireylerin koşudaki hız performansları ölçülmek isteniyor. Ancak siyahı bireyler için kullanılan kronometre "yavaş" çalışmaktadır. Yapılan ölçme sonucundaki (siyah ve beyaz bireylerin oluşturduğu) gruplar içindeki performans sıralaması görece olarak doğrudur. Fakat grupların en iyi performans gösteren bireylerinin karşılaştırılmaları ya da her iki grubun ortalama performanslarının karşılaştırılmaları için ölçme aracı yanlı olarak nitelendirilir."

Örnek 2: Osterlind (1983)

Okulun atletizm takımı antremanlarını kasabanın caddelerinde yapmaktadır. Okulda yapılan bir testte öğrencilerin kasaba bilgisi test edilmek isteniyor. Yapılan bu ölçme atletizm antremanlarına katılan öğrencilerin lehine işlemiştir.

Örnek 3: Camilli & Shepard (1994):

Dekoy¹: Ördek

- a) Ağ: Kelebek
- b) Örümcek Ağ: Örümcek
- c) Tuzak Yemi: Balık
- d) Patika: Kısayol

Decoy: Duck

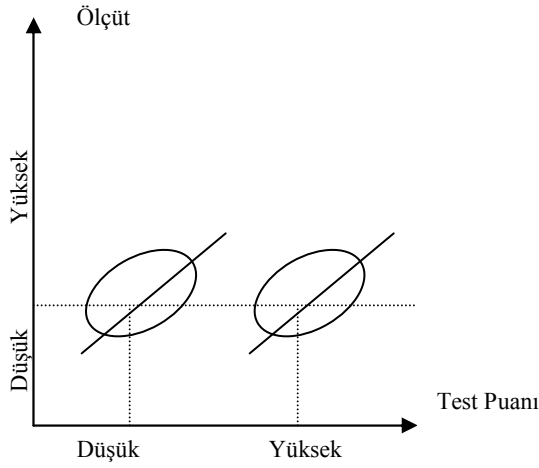
- a) Net: Butterfly
- b) Web: Spider
- c) Lure: Fish
- d) Detour: Shortcut

Bu üç sorudan birincisi ölçme aracı hatalı olduğu için ölçme sonuçları sistematik hata içermektedir ve yapılan ölçmede yanlılık vardır (bilgi kuramına göre bilgi bozulumu vardır). İkinci örnekte ölçme, atletizm takımında olanlara göre "yanlı" davrandığı düşünülebilir, ancak testi alan öğrencilerin bilgi ve tecrübeleri ölçme aracının alt gruplara göre yanlı davrandığı anlamına gelmemektedir ($X \rightarrow T$ kanalında bilgi bozulumu söz konusu değil). Burada madde etkisinden sözedilebilir.

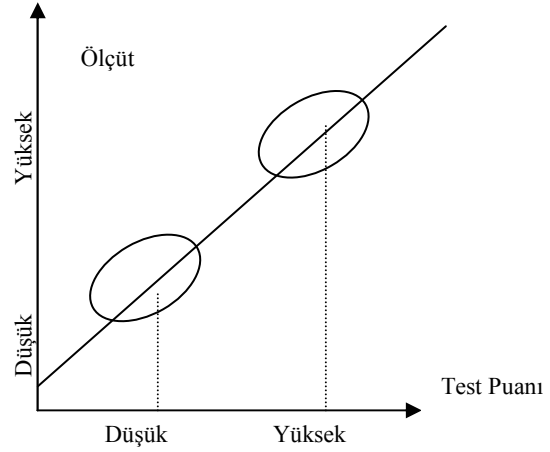
¹ Decoy; ördek avında ördekleri tuzağa düşürmek için kullanılan bir çeşit tuzak düdüğü.

Üçüncü örnek ise erkek öğrencilerin kız öğrencilere göre daha kolay yanıtlayabildiđi bir test sorusudur. Burada ölçülmek istenen yapı “ilişki kurma” üzerinedir ama kullanılan literatür “balık avcılığı” ile ilgili olduđu için kız öğrenciler için Camille & Shepard (1994) ise bu maddeyi “yanlı” madde olarak nitelemektedir.

Camille & Shepard (1994) yanlılığı ortaya çıkarmaya yönelik yöntemleri dışsal ve içsel yöntemler olarak vermişlerdir. Dışsal yöntemlere göre ölçüt ve test puanları arasında alt gruplara göre farklı regresyon çizgileri elde edilebiliyorsa ilgili madde yanlıdır (Şekil 3). Eğer ölçüt ve test puanları arasındaki regresyon çizgisi alt gruplara rağmen sürekli ise ilgili madde yansızdır (Şekil 4).



Sekil 3: Yanlı bir madde



Sekil 4: Yansız bir madde

Dışsal yan çözümlenmeleri için gösterge deđişkenli regresyon modeli (regression with dummy variable) kurulabilir.

$$\text{Ölçüt} = \text{Test Puanı} + \text{Grup} + \text{Hata} \quad \text{Grup} = 0,1$$

Böylelikle genel model için grupların etkisi araştırılabilir.

İçsel yöntem, ilk zamanlar dışsal bir ölçüt bulmanın zor olduđu durumlarda test puanlarının içsel ölçüt olarak kullanılıyla geliştirilmiştir. Buna göre, ölçülmek istenen T özelliđine göre gruplar arasında bir farklılık söz konusu olduđunda bu farklılık, yanlılıktan kaynaklanacağı gibi performanstaki gerçek farklılıktan da kaynaklanabilir. Bu durumda,

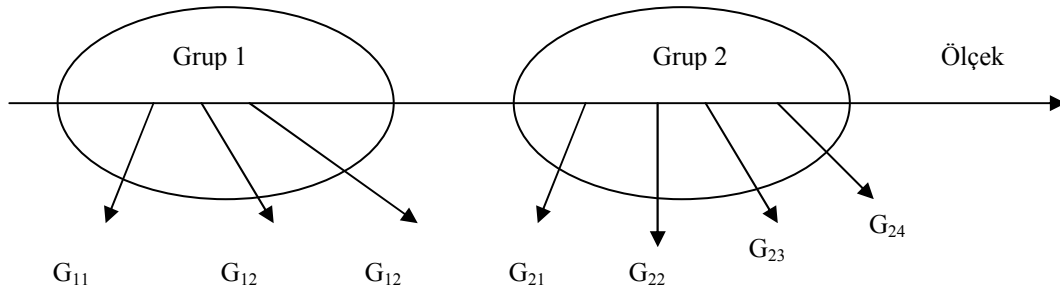
ölçmeye konu olan özellik göz önüne alınarak madde etkisi ya da yanlılık tanımının yapılması gerekmektedir.

Bu aşamada bazı kavramların karşılıklarına yer vermek gereklidir:

Madde Etkisi (Item Impact): Gruplararasıda varolan gerçek farklılıktır. Farklı gruplardaki yanıtlayıcıların bir maddeyi doğru yanıtlama/uygun bulma² olasılıklarındaki farklılıkları ifade eder. Bu farklılıklar bir maddenin ölçmeye çalıştığı belli bir yetenek düzeyinde gruplar arasında var olan gerçek farklılıktır.

Madde Yanlılığı (Item Bias): Testin amacına uygun olmayan test koşullarından ya da test maddelerinin bazı karakteristik özelliklerinden dolayı bir maddeyi bir grubun doğru yanıtlamasının diğer gruba göre daha az/çok olması durumudur.

Diferansiyel Madde Fonksiyonu: Maddenin ölçmek istediği yetenek düzeyinde yapılacak bir karşılaştırmada farklı gruplardaki yanıtlayıcıların ilgili maddeyi doğru yanıtlama olasılıklarındaki farklılıkları ifade eder (Zumbo,1999).



Şekil 5: Grupların ölçek üzerindeki konumları

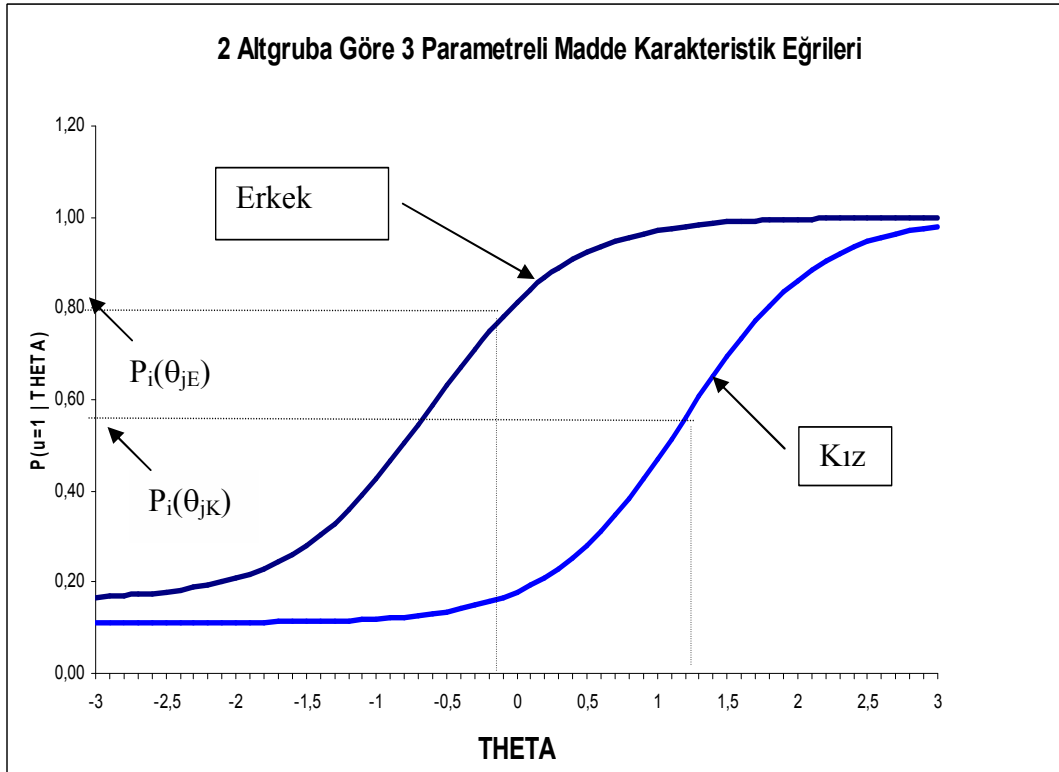
Yukarıdaki ölçekte gruplar farklı boyutta yer almışlardır. Bu yerleşim, ölçek etkisinden (madde etkisi) kaynaklanabileceği gibi, ölçeğin yanlı olmasından da ya da hem madde etkisinden hem de yanlılıktan kaynaklanabilir.

² “Maddeyi doğru yanıtlama” bilişsel yapılara ilişkin bir deyimdir. Duyuşsal yapılara ilişkin “maddeyi uygun bulma” deyimini kullanılmaktadır. Bu çalışmada verilen örnekler ya da benzetmeler için bilişsel yapıya ilişkin kavramlar kullanılacaktır.

Diferansiyel Madde Fonksiyonu (Differential Item Function)

DMF aynı yeteneğe sahip ancak farklı gruplarda yer alan bireylerin ilgili maddeyi doğru yanıtlama olasılıklarındaki farklılığı işaret etmektedir. Bu farklılık, öğrencilerin yetenek, test puanları gibi içsel bir ölçüt ile karşılaştırılarak ortaya çıkarılmaktadır.

Zumbo'ya göre (1999), yanlılığın varlığı iki yöntem ile araştırılabilir; bunlar sırasıyla yargı ve istatistiksel yöntemlerdir. Yargı yöntemi uzman kanısına dayanmaktadır. İstatistiksel yöntemler ise daha objektiftir.



Şekil 6: Alt gruplara Göre Madde Karakteristik Eğrileri

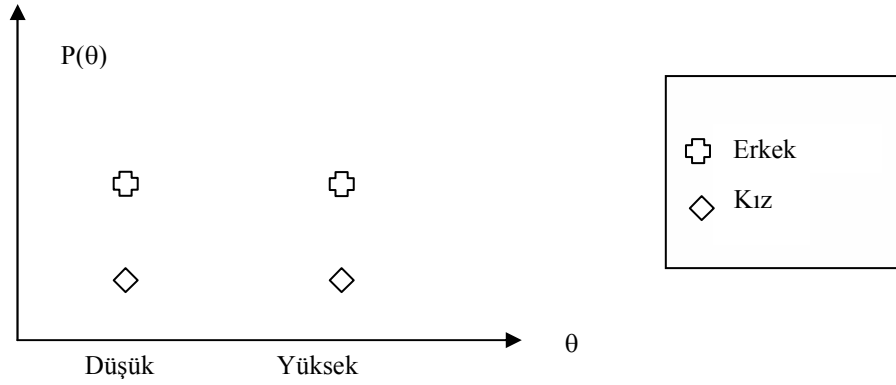
Şekil 6'da, bir maddenin cinsiyet alt grubuna ilişkin DMF verilmiştir. Örneğin; $\theta=0$ düzeyinde alt grupların maddeyi doğru yanıtlama olasılıkları değişik değerler almaktadır. Burada bir fonksiyonel bozulmuş³ söz konusudur. Bu nedenle tek bir madde fonksiyonundan (madde karakteristik fonksiyonu – item characteristic function/curve) söz etmek olanaksızdır ve bu durum diferansiyel madde fonksiyonunun (DMF) ile açıklanabilir.

Diferansiyel madde fonksiyonu TekBiçimli (uniform) DMF (TB-DMF) ve TekBiçimli Olmayan (non-uniform) DMF (TBO-DMF) olmak üzere iki türdür.

³ Matematiksel fonksiyonlarda, değişkenin fonksiyonda yalnızca bir görüntüsü olabilmektedir. Bir değişkenin (herhangi bir değeri) fonksiyonda birden fazla görüntüsü oluşabiliyorsa bu durum tek bir fonksiyon ile açıklanamaz.

Tekbiçimli DMF (TB-DMF):

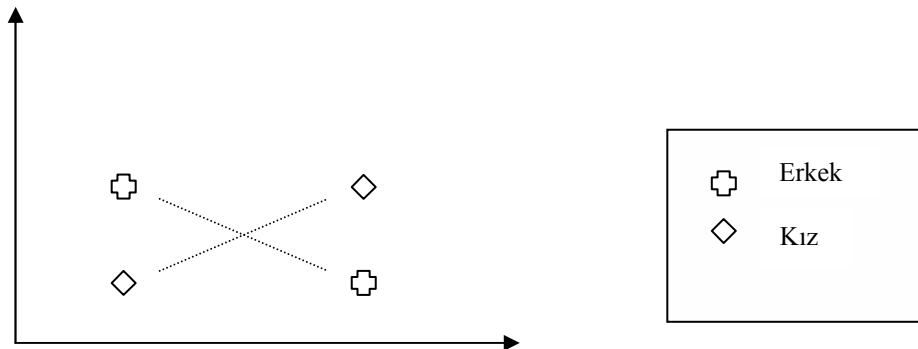
Gruplar arasında ilgili maddeyi dođru yanıtlayma olasılıđındaki farklılık, bütün yetenek düzeylerinde tutarlılık göstermesi durumunu belirtir.



Şekil 7: Erkek ve Kız Bireylerin, Düşük-Yüksek Yetenek Düzeylerindeki Maddeyi Dođru Yanıtlayma Olasılıkları

Örneđin altgrup olarak cinsiyet alındıđında, Şekil 6'da görüldüğü gibi, erkeklerin maddeyi dođru yanıtlayma olasılıđı farklı yetenek düzeylerinin her noktasında kızlara göre daha yüksek ise, yani düzlemde bir kesişim sözkonusu deđilse bu şekildeki DMF'na tekbiçimli DMF adı verilmektedir.

Tekbiçimli Olmayan DMF (TBO-DMF):



Şekil 8: Erkek ve Kız Bireylerin, Düşük-Yüksek Yetenek Düzeylerindeki Maddeyi Dođru Yanıtlayma Olasılıkları

Gruplar arasında varolan, maddeyi dođru yanıtlayma olasılıđındaki farklılık, bütün yetenek düzeylerinde sabit deđilse bu tür DMF'na tekbiçimli olmayan DMF adı verilmektedir. Bir başka ifade ile; yetenek düzeyleri ile grup kategorileri arasında bir kesişim var ise TBO-DMF olduđu söylenebilir.

Camilli ve Shepard (1994:59), TB-DMF'na aynı zamanda "tutarlı DMF" adı vermekteler ve tanımını ise; a parametresi sabit iken b parametresindeki tutarlı farklılık olarak yapmaktalar. Aynı şekilde, TBO-DMF için "tutarsız DMF" olarak nitelemekteler ve b parametresi aynı iken a parametresindeki değişim olarak ifade etmekteler.

Fidalgo, Mellenberg ve Muniz (2000), alt gruplar için elde edilen madde ayırıcılık parametresindeki farklılıkların yarattığı DMF'na simetrik TBO-DMF, hem a hemde b parametrelerinde farklılık söz konusu ise bu durumda ortaya çıkan DMF için ise asimetrik TBO-DMF adını vermişlerdir.

EK:

Bilgi Kuramı Üzerine Genel Bilgiler

Eğitim ve psikoloji alanında yapılan ölçme işlemleri fiziksel ölçmelere göre daha titiz çalışmalar gerektirmektedir. Bunun nedeni, ölçülecek olan özellikten ve ölçmede kullanılacak olan araçtan kaynaklanmaktadır. Ölçmeye konu olan özelliğin bir davranış ya da psikometrik bir değişken olması, bu tür ölçmelerin psikofizik boyutta ele alınmasına neden olmaktadır. Bu ise davranış bilimlerinin bir özelliğidir.

Psikofizik ölçmeler, ölçülmek istenilen özelliğin gerçek büyüklüğü ile ölçmecinin algıladığı büyüklük arasındaki bağıntı üzerinde durur. Bu bağıntının fonksiyonel gösterimini Fechner,

$$\text{Algılanan Büyüklük} = f(\text{Gerçek Büyüklük})$$

şeklinde bir bağıntı olduğunu ve bu bağıntının açılımının ise;

$$\text{Algılanan Büyüklük} = C \cdot \text{Log}(\text{Gerçek Büyüklük})$$

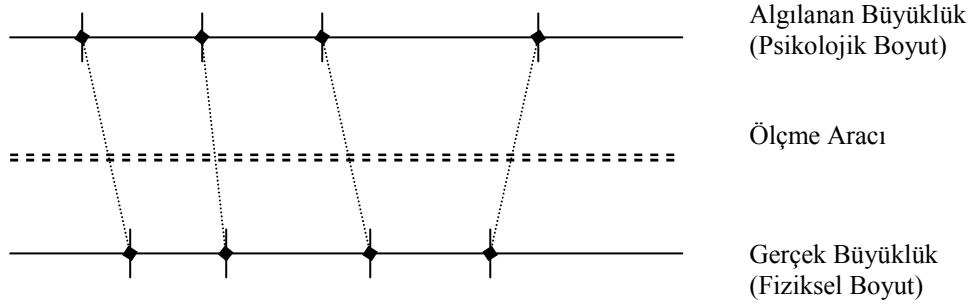
olduğunu belirtmiştir (Turgut ve Baykul, 1992:11).

Bu yaklaşım entropi kavramı ile açıklanabilmektedir. Entropi kavramı, bilgi kuramı (information theory) ve belirsizlik kuramı (uncertainty theory) ile belirtilen yaklaşımlarda kullanılan bir kavramdır ve bir düzlemdeki belirsizlik (uncertainty) miktarını verir. Entropinin tersi ise düzlemdeki belirsizlikten arındırılmış olan "bilgi miktarı" olarak adlandırılmaktadır. Bu yöntem Fischer tarafından istatistik alanında uygulanmış ve örneklem (bilinen) ile evren (bilinmeyen) ya da model-sistem arasındaki bağıntıyı verecek şekilde düzenlenerek bugün yaygınca kullanılan test istatistikleri geliştirilmiştir.

Psikofizik ölçmelerde, algılanan büyüklük ile gözlenen büyüklük arasındaki bağıntılar da entropi ölçüsü ile açıklanabilmektedir. Örneğin Woodcock yetenek ölçeği (Hambleton & Swaminathan, 1985:59) entropi üzerine kurulu bir standartlaştırma ölçeğidir ve Fechner'in bağıntısı ile benzerlik taşımaktadır.

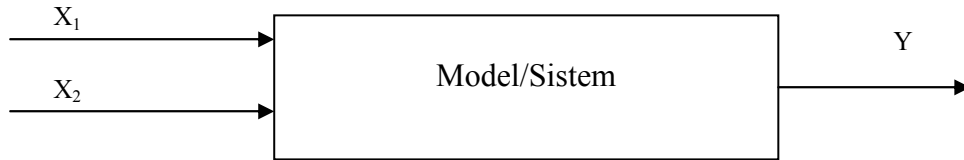
$$W_0 = C_1 \log Z + C_2$$

Aşağıda, psikofizik uzayda yapılan ölçmelerin şematik yapısı verilmiştir.



Şekil 1: Psikofizik ölçme durumları

Şekil 1'den de görülebileceği gibi; hatalardan arındırılmış bir ölçme yapmak olanaklı olsa, algılanan büyüklük ile gerçek büyüklük eşit olacaktır. Ancak burada ortaya çıkabilecek iki sorun vardır. Bunlar sırasıyla, algılanan ile gerçek büyüklük arasında bağıntının doğru bir şekilde kurulabilmesi (entropi bağıntısı) ve ortaya çıkacak olan hatanın enküçüklenmesi (minimized) olarak belirtilebilir.



X_1 'in sisteme girme p_1 olasılığı taşıdığı varsayılır ise sistem için X_1 'in sağladığı bilgi miktarı, Shannon tarafından geliştirilen ve Shannon Entropisi olarak bilinen bağıntı ile açıklanmaktadır (Yurdugül, 1997).

$$I(X) = \log\left(\frac{1}{P_1}\right)$$

$$I(X) = -\log P_1$$

Sistem bir bütün olarak düşünülürse bu durumda ortalama bilgi miktarına (Entropi) ihtiyaç vardır:

$$H(X) = \sum_i P_i \log\left(\frac{1}{P_i}\right)$$

$$H(X) = -\sum_i P_i \log P_i$$

Eğer X değişkeninin dağılımı sürekli ise;

$$H(X) = \iint f(x, y) \log\left(\frac{1}{f(x)f(y)}\right) dx dy = -\iint f(x, y) \log[f(x)f(y)] dx dy$$

Pearson, bu ifadede x ve y'nin normal dağılıma sahip olduğu varsayımı altında karşılıklı bilgi (mutually information) yaklaşımı ile Pearson çarpım momentler korelasyon katsayısını geliştirmiştir (Yurdugül, 1997).

$$I(X) = -\frac{1}{2} \log(1 - \rho^2)$$

Bilgi kuramındaki bağıntıda yer alan logaritmada taban olarak herhangi bir sayı kullanılabilir; ancak en çok kullanılan tabanlar: 2, e ve 10'dur. Eğer taban 2 alınırsa bu durumda bilginin birimi "Bit" olur.

Örnek: Bir paranın atılması deneyi ele alındığında X rastlantı değişkeni "tura" gelme olayını temsil etmek üzere.

$$P(X_1)=1/2 \quad Q=1-P(X_1)=1/2$$

$$H(X_1) = \sum_{i=1}^2 P_i \log_2 P_i = (1/2) \log_2 2 + (1/2) \log_2 2 = 1 \quad (\text{bit})$$

$H(X_1)$, bileşen (marginal) entropi fonksiyonudur. Birleşik (jointly) entropi fonksiyonu ise;

$$H(X_1 X_2) = -\sum_i \sum_j p_{ij} \log p_{ij}$$

Eğer X sürekli bir dağılıma sahip ise bu seferde $H(X_1 X_2)$ ifadesi aşağıdaki şekilde tanımlanır:

$$H(X_1 X_2) = -\iint f(x_1, x_2) \log\left(\frac{f(x_1, x_2)}{f(x_1)f(x_2)}\right) dx_1 dx_2$$

Bu ifade ve ek olarak entropi ifadesi Fischer ve Pearson istatistiklerinin temelini oluşturur (Joe, 1989).

$$I(X_1;Y)=H(X_1)+H(Y)-H(X_1Y)$$

Bu ifade, bilgi kanalındaki ortak bilgi miktarı olarak adlandırılmaktadır. Aynı şekilde alınan mesajın iletilen mesaja ilişkin taşıdığı bilgi miktarı olarak da adlandırılmaktadır.

1957 yılında Khinchin ve 1969 yılında Eppler, bu ifadeyi "doğru iletilen bilgi" olarak adlandırılmaktadır. Eppler, bilgi kanalının etkinlik derecesi olarak aşağıdaki eşitliğini önermiştir (Esensoy, 1984).

$$\eta = \frac{I(A;B)}{H(A)} = \frac{H(A) - H(A|B)}{H(A)}$$

$$\eta = 1 - \frac{H(A|B)}{H(A)}$$

Baykal (1994), bilgi kanalının etkinlik derecesini eğitimde ölçme alanına uyarlayarak, aşağıda gösterilen (g) yapısal geçerlik göstergesi olarak ifade etmiştir.

$$g = \frac{I(X;Y)}{H(X,Y)}$$

Test maddelerinin ortaya koyduğu belirsizlik miktarını ise;

$$H(X, Y) = \log K - \frac{1}{K} \sum \sum N \log N(X, Y)$$

şeklinde belirtmiştir.

Baykal (1980-81), benzer bağıntılardan yola çıkarak yaptığı çalışmada; hatanın ölçüsü olarak, bilgi kuramından alınan belirsizlik bağıntılarından yararlanarak "öğrenme, başarı ve erişime bakarak yapılacak değerlendirmeler denek sayısından ve deneklerarası değişkenlikten bağımsız olarak yürütülebilir" ifadesini kullanmıştır.

Entropinin Özellikleri

- Olasılığın sürekli bir fonksiyonudur.
- Eşit olasılıklı olaylar için maksimumdur.
- Olay sayısı (deney sonucu sayısı) arttıkça entropy de artar.
- $H(X_1, X_2) \leq H(X_1) + H(X_2)$

KAYNAKÇA

- BAYKAL Ali, "Davranış Ölçümünde Yapısal Geçerlik Göstergesi", Türk Psikoloji Dergisi, 1994, 9 (33), 45-50
- BAYKAL Ali, 1980-81, "Davranış Ölçümünde Hata Göstergesi: Entropi", Boğaziçi Üniversitesi, Eğitim Dergisi, VIII-IX
- BAYKUL, Yaşar. *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*, ÖSYM Yayınları, Ankara, 2000
- BEERS, Yardley. *Introduction to The Theory of Error*. Addison-Wesley, Colorado, 1957.
- CAMILLI, Gregory., SHEPARD Lorrie A., *Methods for Identifying Biased Test Items*, Volume 4, SAGE Publications, 1994.
- DUDEWICZ, E. J. ve MISHRA, S. N., *Modern Mathematical Statistics*, John Wiley, 1988.
- ESENSOY, Ömer. "Çapraz Tabloların Entropi İlkesi ile Değerlendirilmesi", E.Ü. Mühendislik Fakültesi Dergisi, Seri:E Uygulamalı İstatistik, Cilt 2 Sayı:1 1984.
- HAMDAN, M. A., TSOKOS, C., P., "An Information Measure of Association in Contingency Tables", *Information and Control*, vol:19, 174-179, 1971.
- HAMBLETON, R. K. ve H. SWAMINATHAN. *Item Response Theory: Principles and Application*. Kluwer-Nijhoff Publishing, Boston, 1985
- JOE, H. 1989, "Relative Entropy Measures of Multivariate Dependence", *Journal of the American Stat. Assoc.* 84, 157
- OSTERLIND, J.S. *Test Item Bias*. London: Sage Publications, 1983
- TURGUT, M. F. *Eğitimde Ölçme ve Değerlendirme Metotları*, Saydam Matbaacılık, Ankara, 1992
- TURGUT, M. F. ve Y. BAYKUL. *Ölçekleme Teknikleri*, ÖSYM Yay., Ankara, 1992
- YURDUGÜL, Halil. Kategorik veriler için bağımlılık ve ilişki ölçülerinin karşılaştırılması, Yayınlanmamış Bilim Uzmanlığı Tezi, H.Ü. Fen Bilimleri Enstitüsü, 1997
- ZUMBO, B. D. 1999 *A Handbook on the Theory and Methods of Differential Item Functioning: Logistic Regression Modelling as a Unitary Framework for Binary and Likert-Type Item Scores*. Ottawa, Directorate of Human Resources Research and Evaluation, Department of National Defense.