

Ölçeklerin değerlendirilmesinde niçin Rasch analizi kullanılmalıdır?

Why is it necessary to use Rasch analysis when evaluating measures?

Atilla Halil Elhan, Yıldır Atakurt

Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı

Rasch analizi herhangi bir kişinin herhangi bir maddedeki bir kategoriye seçme olasılığının kişi yetenek düzeyi ile madde zorluk düzeyi arasındaki farkın bir lojistik fonksiyonu olduğunu varsayar. Sağlık alanında kullanılan anket ve ölçeklerin birçoğu sıralı ölçeğe sahiptir. Bu nedenle, maddelere verilen doğru cevapların toplanmasıyla elde edilen ham puanları kullanarak anket ya da ölçek değerlendirilmeye çalışıldığı zaman birtakım sorunlarla karşılaşılır. Rasch analizi bu sorunların üstesinden gelmek için kullanılan yöntemlerden biridir.

Anahtar sözcükler: *Geçerlilik, güvenilirlik, lojit, Rasch model*

Rasch analysis assumes that the probability that a person will affirm an item or category within an item is a logistic function of the difference between the person's ability and the difficulty of the item, and only a function of that difference. Most of the questionnaires and measures have ordinal scales. For this reason, when it is tried to evaluate them by using raw scores, there will be some problems. Rasch analysis is one of the methods which copes with problems.

Key words: *Reliability, validity, logit, Rasch analysis*

Özellikle psikoloji ve rehabilitasyon alanlarında uygun tedavinin belirlenmesi, izlenmesi ve etkinliğinin değerlendirilmesi amacıyla çeşitli sonuç değerlendirim ölçekleri kullanılmaktadır. Bu ölçeklerin sadece klinik uygulamalarda değil, özellikle gelişmiş ülkelerde daha kapsamlı olarak, toplum içi ve toplumlar arası farklı tedavi programlarının karşılaştırılması, sağlık politikalarının belirlenmesi ve sağlık hizmetleri için kaynak tahsisi gibi alanlarda da kullanımı gündeme gelmiştir. Sonuç değerlendirim ve izleniminin en doğru şekilde yapılabilmesi için, bu ölçeklerin kullanıldıkları toplumlara göre din, dil ve sosyo-kültürel adaptasyonlarının yapılması, geçerlilik ve güvenilirliklerinin gösterilmesi şarttır. Gerek toplum içinde aynı hastalık grubunda farklı tedavi programlarının ve farklı hasta gruplarının sonuçlarının karşılaştırılması, gerekse toplumlar arası karşılaştırmalar yapılabilmesi için bu ölçeklerin ulusal ve uluslararası düzeyde standart hale getirilmesi gerekmektedir. Bu standartlaştırma işleminde kullanılan klasik psikometrik yöntemlerin yeterli olmadığı öne sürülmüş ve Rasch analizi gündeme gelmiştir.

Sağlık alanındaki sonuç ölçümlerin birçoğu sıralı ölçeğe sahiptir. Bu nedenle sağlık bakım sonuçlarını değerlendirmek için aritmetik işlemlerin ve parametrik istatistiksel yöntemlerin kullanılması mümkün olmamaktadır (1, 2). Değişkenin sıralı ölçeğe sahip olduğu durumlarda, sadece hastaların fonksiyonel durumlarında değişiklik olup olmadığı test edilebilmekte, eğer değişiklik varsa bu değişikliğin miktarı belirlenmemektedir. Yukarıda belirtildiği gibi sıralı değişkenlere ilişkin sayısal analizlerin yetersizliği bilinmektedir (3, 4). Sıralı ölçekli verilerin birçoğundan daha güçlü sonuçlar veren aralık ölçekli ölçümler elde edilebileceği

4. İstatistik Günleri Sempozyumu'nda (20-21 Mayıs 2004 Kuşadası, İzmir) poster olarak sunulmuştur.

Geliş tarihi: 27 Aralık 2004 • Kabul tarihi: 25 Ocak 2005

Yazışma adresi:

Atilla Halil Elhan
Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı,
Ankara

Tel. : +90 312 310 30 10/396

Faks : +90 312 310 64 94

E-posta : ahelhan@yahoo.com

de bir gerçektir (5, 6). Bununla beraber, sıralı ölçümlerden başarılı bir şekilde aralık ölçümlerinin elde edilmesi bir takım ileri düzey araştırmalar gerektirmektedir. Genellikle anketlerden elde edilen toplam puanlar maddelere verilen doğru cevapların ya da başarı sayılarının toplanmasıyla elde edilir (7). Bu tür değişkenler de sıralı ölçeklidir.

Sağlık bakımının etkinliğini belirleme oldukça önemli bir sorundur. Rasch analizi özellikle fiziksel tıp ve rehabilitasyon dalında sıralı ölçekli ölçümleri aralık ölçekli hale dönüştürmek ve sonuçta sağlık bakımının etkinliğini belirlemek amacıyla oldukça yaygın bir şekilde kullanılmaktadır. Rasch analizinin sağlık bakımındaki uygulamaları klinik sonuçların ölçümlerine yeni bir bakış açısı getirmiştir (1).

Ölçümlerin aralık düzeyinde yapıldığı değişkenlerde sayılar arasındaki aralıklar eşittir. Bu nedenle aralık düzeyli bir değişkende ölçek ile cevap arasında aritmetik bir ilişki vardır (2, 8). Örnek olarak, 40 yaşındaki bir kişi, 20 yaşındaki bir kişiye göre 2 kat yaşlıdır ya da sürekli ölçeğe sahip bir yürüme özür ölçeğinde (ki bu da yürüme mesafesi olabilir) eğer puanın artması özürölülük derecesini arttırıyorsa, 12 puanına sahip bir kişi, 6 puanına sahip başka bir kişiye göre iki kat daha fazla özürölüdür denilebilir. Diğer taraftan, yürüme özürölülüğünü ölçen sıralı bir değişken ise yürüme özür ölçeğinde 12 puana sahip bir kişinin, 6 puana sahip bir kişiye göre daha fazla özürölü olması yorumundan başka bir şey söylenemez. Bu nedenle değişkenlerin ölçüm düzeyinin aralık mı? yoksa sıralı mı? olduğuna karar vermek önemlidir. Aksi takdirde hangi istatistiksel yöntemin kullanılmasının uygun olacağı konusunda birşey söylemek mümkün değildir. Bununla beraber, parametrik istatistiklerin kullanılması için değişkenin sadece normal dağılıma sahip olması değil, aynı zamanda ölçümün aralık düzeyli olması gerektiği de bir gerçektir (8).

Rasch analizi, kişi yeteneği ile madde zorluk değerlerini ortak bir eksen boyunca yerleştirmeye çalışır. Verilen bir yetenek düzeyindeki bir kişinin gerçekleştirilmesi istenen göreve karşı ne yapabileceği olasılığı kestirilir.

Bu çalışmada, ölçeklerin değerlendirilmesinde karşılaşılan sorunlarla başa çıkmak için neden Rasch analizinin kullanılmasının gerektiği açıklanacaktır.

Maddelere verilen doğru cevapların toplanmasıyla elde edilen ham puanları kullanarak herhangi bir anket ya da test değerlendirilmeye çalışıldığı zaman karşılaşılan bazı sorunlar aşağıda verilmiştir:

i. Anket veya testlerde kullanılan kategoriler arasındaki farkların eşit olmaması

Anketler, değerlendirme testleri ve psikolojik testlerin birçoğu sıralama ölçekli yapıya sahiptir. Maddelere verilen cevaplara ilişkin puanların toplanması gibi standart ana-

liz yöntemleri, seçenekler arasındaki aralıkları eşit kabul ederek test işlemini gerçekleştirir. Gerçekte ise seçenekler arasındaki aralıklar her zaman eşit değildir. Birçok araştırma sonucunda seçenekler arasındaki farkların eşit olmadığı gösterilmiştir. Örnek olarak, romatoid artritli bir hastaya aşağıdaki sorunun sorulduğu varsayalım.

Soru: Beş basamak merdiven çıkabiliyor musunuz?

- a) Hiç zorluk çekmeden yapıyorum (0 puan)
- b) Biraz zorlukla yapıyorum (1 puan)
- c) Çok zorlukla yapıyorum (2 puan)
- d) Hiç yapamıyorum (3 puan)

Gerçekte romatoid artritli bir hastanın bu sorudan 0 puan yerine 1 puan, 1 puan yerine 2 puan, 2 puan yerine 3 puan alma zorluk düzeyleri (olasılıkları) arasında fark vardır. Anket ya da testleri değerlendirme aşamasında ham puanları kullanmak kategoriler arasında aşağıdaki gibi eşit farklılık olduğunu kabul etmek demektir.



Bunun yanı sıra, Rasch analiziyle kategoriler arasında geçiş zorluk düzeyleri hesaplandığında, aşağıdaki gibi kategoriler arasında eşit olmayan farklılığın olduğu görülmektedir. Bu nedenle, kategori puanlarının (ham puanların) toplanması, ortalamasının alınması gibi işlemlerin yapılması doğru olmamaktadır (2, 9).

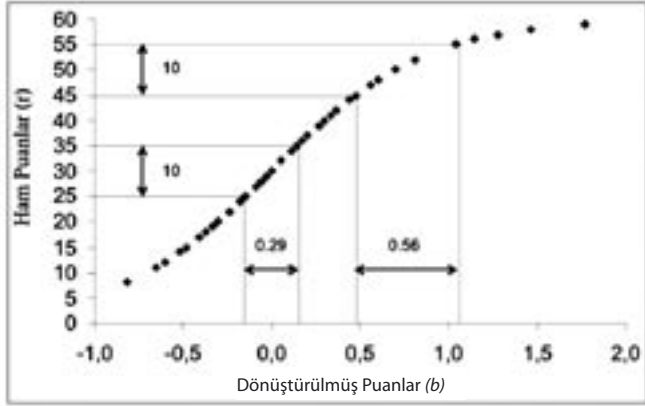


ii. Maddelerin hepsinin eşit zorlukta olmaması

Tutum, beceri, bilgi düzeyi ve özürölülük derecesi gibi kavramlar anket ve testlerle ölçülürken, bu anket ya da testlerdeki maddelerin hepsinin ölçek üzerindeki yerleşimi aynı nokta üzerinde değildir. Başka bir ifadeyle, bazı maddelerin diğerlerinden daha yoğun bilgi ve/veya beceri gerektirmeleri nedeniyle maddelerin hepsi eşit zorlukta değildir. Bununla beraber, maddelerin hepsinin ölçek üzerinde aynı noktada bulunması, ölçülmek istenen değişken hakkında ek bir bilgi sağlanmamasına neden olmaktadır. Maddelerin eşit zorlukta olmaması nedeniyle bu maddelerden elde edilen ham puanlardan yola çıkarak yorum yapılması yanıltıcı olabilmektedir.

iii. Kayıp verilerle başa çıkamama

Anket ya da testlerden ham puanlar hesaplanırken cevap verilmeyen soru ya da maddelere ne yapılacağı sorun olmaktadır. Örnek olarak 2 hastanın 8 maddelik sıralama ölçekli bir ankete verdiği cevaplar aşağıdaki gibi olsun.



Şekil 1: Hastaların ham puanlarına karşı dönüştürülmüş puanların saçılım grafiği.

Madde	1	2	3	4	5	6	7	8	Ham Puan
Hasta A:	2	3	3	4	12
Hasta B:	1	2	2	2	2	2	2	3	16

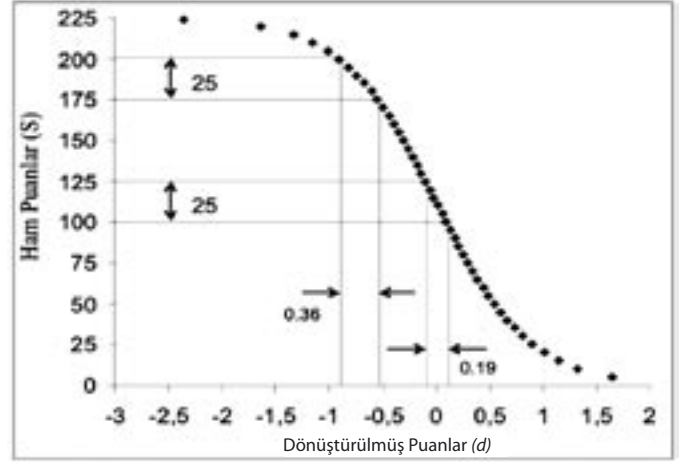
Hangi hasta daha yeteneklidir? Hasta B'nin ham puanı, hasta A'nın ham puanından daha yüksek olmasına rağmen, ortak cevap verdikleri 4 madde göz önünde bulundurulduğunda durum tam tersi olmaktadır. Bu nedenle kayıp veriler söz konusu olduğu zaman ham puanları kullanmak araştırmacıları yanlış sonuçlara götürebilmektedir.

Madde	1	2	3	4	5	6	7	8	Ham Puan
Hasta A:	2	3	3	4	12
Hasta B:	1	2	2	3	8

Rasch analizinde kayıp veriler sorun yaratmamaktadır. Çünkü, Rasch analizi bir defada sadece bir gözlem için işlem yapar. Yani her gözlem (x_{ni}) için beklenen değeri ayrı ayrı hesaplar. Bu hesaplama işlemi sırasında kayıp veriler atlanır. Çalışmadaki her kişi için gözlenen değerlerin toplamıyla beklenen değerlerin toplamının karşılaştırılmasından kişi yetenek düzeyi hesaplanır. Benzer şekilde her madde için de gözlenen değerlerin toplamıyla beklenen değerlerin toplamının karşılaştırılmasından madde zorluk düzeyleri hesaplanır. Bu toplama işlemi sadece değeri bilinen veriler üzerinden yapıldığından kayıp veriler için herhangi bir düzeltme, ayarlama yapılması gerekmemektedir.

iv. Maddelere verilen beklenmedik cevapların belirlenememesi

Aşağıda verilmiş olan sıralama ölçekli anketteki 8 maddenin en kolaydan en zora doğru sıralandığı varsayalım. Ham puanlarla yapılan analizlerde 5. maddeye verilen bek-



Şekil 2: Maddelerin ham puanlarına karşı dönüştürülmüş puanların saçılım grafiği.

lenmedik cevaplar göz ardı edilmiş olmaktadır. Rasch analizinde madde zorluk düzeyleri ve kişi yetenek düzeyleri için gerçekleştirilen uyum testleri (INFIT ve OUTFIT) sonucunda bu tür beklenmedik cevaplar belirlenmektedir.

En kolay → En Zor

Madde	1	2	3	4	5	6	7	8	Ham Puan
Hasta A:	7	7	6	6	(1)	5	5	4	41
Hasta B:	4	4	3	3	(7)	1	1	1	24

v. Örneklemden bağımsız madde zorluk düzeylerinin ve testten bağımsız kişi yetenek düzeylerinin kalibrasyon gerekliliği

Anketten bazı maddeler çıkarıldığında kişilerin ham puanları azalacak, ya da zor maddelerden birkaçı daha kolay olan maddelerle değiştirilirse kişilerin puanları yükselecektir. Bu sorunun nedeni, kişilerin puanlarının anketteki madde sayısına ve bu maddelerin zorluk derecelerine bağlı olmasıdır. Bu yüzden farklı anketlerden elde edilen ham puanların karşılaştırılması mümkün değildir. Farklı çalışmalardan elde edilen kişi ham puanlarını, madde zorluk düzeylerinden arındırarak karşılaştırmak gerekmektedir. Aynı sorun madde zorluk düzeyleri içinde geçerlidir. Bu puanlar, örneklemden kişi sayısı ve bu kişilerin davranışlarından etkilenirler. Bazı kişiler analizden çıkartılırsa, o zaman maddelerin ham puanları düşecektir. Aynı amaçla geliştirilmiş farklı anketlere ait madde ham puanlarını karşılaştırmadan önce ayarlama yaparak ankete katılan kişilerin etkilerini yok etmek gerekmektedir. Rasch analizinde örneklemden bağımsız madde zorluk düzeyleri ve testten bağımsız kişi yetenek düzeyleri kestirilmektedir.

vi. Ham puanların doğrusal ölçek üzerinde ifade edilmiş olmaması

İncelenen değişkeni ham puanları kullanarak yorumlamaya çalışmak kişinin kendisini çarpık bir aynada görmesi gibi birşeydir. Bunun nedeni, değişkenin başında ve sonunda kişilerin ham puanlarının çarpık durumda olmasıdır (Şekil 1). Bu sorunu çözmek için Rasch analizinde madde zorluk düzeyleri lojit metrik sisteme dönüştürülür.

vii. Kişi ve madde puanları için ortak ölçek seçiminin gerekliliği

Şekil 1'de ham puandaki eşit farklılıkların lojit ölçekte eşit olmayan farklılıklara karşılık geldiği görülmektedir. Ham puanlar ele alındığında, 55 ile 45 arasındaki farkla, 35 ile 25 arasındaki fark birbirine eşittir. Fakat bu değerlere dönüşüm uygulandığı zaman 55 ile 45 arasındaki 10 puanlık fark 0.56 lojitlik bir farka denk gelirken, 35 ile 25 arasındaki 10 puanlık fark 0.29 lojitlik bir farka denk gelmektedir. Ham puanlar arasındaki eşit fark, lojit ölçekte yaklaşık birbirinin iki katına ($0.56 / 0.29 \approx 2$) denk gelmektedir. Bu farklılığın nedeni 25 ile 35 arası puanların, hastaların alması daha kolay olan puan aralığına denk gelirken, 45 ile 55 arası puanların 60 limitinin üst sınırına denk gelmesinden kaynaklanmaktadır. Şekil 1'den görüldüğü gibi dönüşüm öncesinde ham puanı yüksek olan hastaların dönüşüm sonrasında da puanları yüksek olmaktadır.

Şekil 2'de verilen eğri Şekil 1'de verilen eğrinin aynadaki görüntüsü gibidir. Madde ham puanları ile onların dönüştürülmüş puanları (lojit) arasındaki ilişki Şekil 2'deki eğri ile tanımlanmıştır.

Kaynaklar

1. Chang WC, Chan C. Rasch analysis for outcomes measures: Some methodological considerations. Arch Phys Med Rehabil 1995;76: 334-339.
2. McHorney CA, Haley SM, Ware JE Jr. Evaluation of the MOS SF-36 physical functioning scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. J Clin Epidemiol 1997;50:451-461.
3. Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. Arch Phys Med Rehabil 1989;70:308-332.
4. Wright BD, Linacre JM. Observations are always ordinal; Measurements, however, must be interval. Arch Phys Med Rehabil 1989;70:857-860.

Burada da ham puandaki eşit farklılıkların lojit ölçekte eşit olmayan farklılıklara denk geldiği görülmektedir. Aşırı uç noktadaki 25 puanlık fark, lojit ölçekte orta aralıktaki 25 puan farka göre daha büyük değişikliğe neden olmaktadır. Yüksek ham puana sahip maddelere dönüşüm uygulandığı zaman düşük lojit puana sahip olurken, düşük ham puana sahip maddelere dönüşüm uygulandığı zaman yüksek lojit puana sahip olmaktadır. Böylece madde puanlarının yönü tersine dönüştürülmüş olmaktadır ve madde puanları artık kişi puanlarıyla aynı yönde hareket etmiş olmaktadır (Şekil 2). Böylece Rasch analiziyle madde ve kişi puanlarını doğrudan karşılaştırmak için aynı değişim aralığı, orijin ve birime sahip olmaları sağlanmış olmaktadır.

Çoğu zaman klasik yöntemlerle anket veya ölçekler değerlendirildiğinde, araştırmanın yapısında bulunan ayrıntılar ve karmaşık yapı yukarıda belirtilen nedenlerden dolayı ortaya çıkarılamamaktadır. Sonuçların ham puan olarak verildiği durumlarda aritmetik işlemler uygulanmadığından, bunlar gerçek ölçümler olmamaktadır. Bu nedenle, kişilerin, maddelerin veya grupların birbiriyle karşılaştırılmasında tutarlı sonuçlar vermez. Madde cevap teorisi altında yer alan yöntemlerden birisi olan Rasch modelini kullanarak ölçeklerin değerlendirilmesiyle yukarıda bahsedilen tüm problemlere çözüm getirilmiş olmaktadır. Bu nedenle, anket veya ölçeklerin değerlendirilmesinde sıralı ölçeğe sahip ham puanlar yerine, Rasch dönüşümlü aralık ölçekli puanların kullanılması ve aritmetik işlemlerin bu puanlar üzerinden yapılması daha doğru olmaktadır.

5. Health and Public Policy Committee, American College of Physicians. Comprehensive functional assessment for elderly patients. Ann Intern Med 1988;109:70-72.
6. Feinstein AR, Josephy BR, Wells CK. Scientific and clinical problems in indexes of functional disability. Ann Intern Med 1988; 105:413-420.
7. Linacre JM, Heinemann AW, Wright BD ve ark. The structure and stability of the Functional Independence Measure. Arch Phys Med Rehabil 1994;75:127-132.
8. Tennant A, Hillman M, Fear J ve ark. Are we making the most of the Stanford Health Assessment Questionnaire? Br J Rheumatol 1996;35:574-578.
9. Tatum DS. Research problems-Rasch solutions. Popular Measurement 1998;1:6-8.