11. BÖLÜM

ÖLÇME SONUÇLARINA UYGULANAN İSTATİSTİKSEL İŞLEMLER

STATISTICAL ANALYSIS OF MEASUREMENT RESULTS

Doç. Dr. Eren Can AYBEK

ÖLÇME SONUÇLARINA UYGULANAN İSTATİSTİKSEL İŞLEMLER

Doç. Dr. Eren Can AYBEK
Pamukkale Üniversitesi

Öz

Bu bölümün amacı ölçme sonuçlarını anlamlandırabilmek için kullanılan test ve madde istatistiklerini açıklamaktır. Bu bağlamda, bir dağılımın hangi noktaya eğilim gösterdiğinin ölçüsü olan merkezi eğilim ölçülerinin yanında dağılımdaki verilerin o merkez noktadan ne kadar yayıldığının göstergesi olarak yayılım ölçüleri aktarılmıştır. Ayrıca farklı ölçme sonuçlarının karşılaştırılmasına olanak sağlayan standart z ve T puanları açıklanmıştır. Test puanları ile hesaplanan bu istatistiklerin yanında testte yer alan maddelerin niteliklerini belirlemek için kullanılan madde istatistik ve parametreleri de yine bu bölümün konuları arasındadır. Bir maddenin, uygulandığı gruba ne kadar zor geldiğinin bir ölçüsü olarak madde güçlüğü; yine bir maddenin ölçmeye çalıştığı özelliğe az ya da çok sahip olan bireyleri ayırmada ne derecede başarılı olduğunun bir ölçüsü olarak madde ayırt edicilik indeksine yer verilmiştir. Bunların yanında yanıtların seçeneklere dağılımı da bölüm kapsamında ele alınan bir diğer konu olmuştur. Madde analizleri öncelikle Klasik Test Kuramı bağlamında ele alınmış olsa da Madde Tepki Kuramı temelinde de madde parametreleri hakkında bilgi verilmiştir. Bölüm, test istatistikleri ve madde analizine dair kuramsal bilgiler verdikten sonra Shiny Item Analysis yazılımı ve örnek bir veri seti kullanarak kuramsal olarak ele alınan bilgilerin uygulamalarını içermektedir. Özellikle Madde Tepki Kuramı için varsayımların test edilmesi, madde – veri uyumunun incelenmesi, madde parametrelerinin kestirimi ve yorumlanması hem istatistiksel değerler hem de grafikler üzerinden yorumlanmış ve açıklanmıştır.

Anahtar Kelimeler

Test İstatistikleri, Madde Analizi, Klasik Test Kuramı, Madde Tepki Kuramı, Shiny Item Analysis

STATISTICAL ANALYSIS OF MEASUREMENT RESULTS

Assoc. Prof. Dr. Eren Can AYBEK
Pamukkale University

Abstract

The aim of this chapter is to explain the test and item statistics used to interpret measurement results. In this context, measures of central tendency, which indicate the point toward which a distribution tends to cluster, are presented alongside measures of variability, which reflect the extent to which the observations in a distribution are dispersed around that central point. Standardized z and T scores, which enable the comparison of different measurement results, are also explained. In addition to these statistics calculated from test scores, the chapter addresses the item statistics and parameters used to determine the characteristics of the items included in a test. Item difficulty is presented as an indicator of how difficult an item is for the group to which it is administered, whereas the item discrimination index is discussed as an indicator of the extent to which an item successfully distinguishes between individuals who possess different levels of the attribute being measured. The distribution of responses across the available options is another topic covered in the chapter. Although item analyses are primarily discussed within the framework of Classical Test Theory, information about item parameters based on Item Response Theory is also provided. After presenting the theoretical foundations of test statistics and item analysis, the chapter demonstrates their practical applications using the Shiny Item Analysis software and a sample dataset. Particular attention is given to testing the assumptions of Item Response Theory, examining item–data fit, and estimating and interpreting item parameters. These processes are explained and interpreted using both statistical results and graphical representations.

Keywords

Test Statistics, Item Analysis, Classical Test Theory, Item Response Theory, Shiny Item Analysis

1. GİRİŞ

Kitabın ilk bölümünde tanımlandığı gibi ölçme, bireylerin eğitime konu olan özelliklerinin gözlenip sonuçların sembollerle, özellikle de sayı sembolleriyle ifade edilmesidir (Turgut, 1997). Bu gözlem sonuçları üzerinde yapılacak istatistiksel işlemler, hem ölçülen bireylerin ölçmeye konu özelliğe ne düzeyde sahip olduğuna hem de kullanılan ölçme aracının niteliğine dair bilgiler sağlar. Örneğin, "Testi alan öğrencilerin başarısı nasıldır? Başarı düzeyleri birbirine benziyor mu? Grubun genelinden aşırı sapma gösteren öğrenciler var mı? Testin 3. maddesi gruba zor mu geldi? Testteki 6. madde ölçülen özelliğe sahip olan ve olmayan öğrencileri iyi bir şekilde ayırabilmiş mi? sorularına yanıt bulabilmek için istatistiksel işlemlere başvurulur.

Kitabın bu bölümünde öncelikle, bir testten alınan puanlar ile testin puanlanabilen en küçük birimi olan maddeden elde edilen puanlar üzerinde yapılan istatistiksel işlemler üzerinde durulmuştur. Daha sonra uygulama başlığı altında söz konusu istatistiksel işlemlerin nasıl gerçekleştirileceği açıklanmıştır. Akabinde uygulanan analizlere ait sonuçların nasıl yorumlanacağına ilişkin örnekler sunulmuştur.

2. TEST PUANLARI ÜZERİNDE YAPILAN İSTATİSTİKSEL İŞLEMLER

Bir test ister uluslararası düzeyde isterse küçük bir sınıfta uygulanmış olsun ölçme sonuçlarını anlamlandırmak için belli istatistiksel değerlere ihtiyaç duyulur. On kişilik bir sınıfın başarı düzeyinin nasıl olduğu ortaya konmak isteniyorsa her öğrencinin başarı puanını ayrı ayrı söylemek yerine tüm sınıfa ait tek bir sayısal değer verilmesi tercih edilir. Örnekteki bir öğrencinin test puanı "veri" ya da "gözlem" iken (Arıcı, 2006), tüm sınıfın puanları bir araya gelerek bir "dağılım" oluşturur (Evans & Ildstad, 2001). Test puanlarından anlam çıkarmak için bireylere ya da maddelere ait verilere ayrı ayrı odaklanmak yerine dağılıma ait özelliklerle ilgilenilir.

Bir testten alınan toplam puanları sıralama düzeyinde mi yoksa eşit aralıklı düzeyde mi kabul etmek gerektiğine yönelik tartışmalar 1940'lı yıllara kadar uzanmaktadır. Crocker ve Algina (2008) bu tartışmaların tarihsel gelişimini ele almış, bu kapsamda farklı araştırmacıların görüşlerinden bahsetmiştir. Stevens (1946) psikolojik testler ile sıralama düzeyinde ölçme yapıldığını, bu nedenle bu testlerden alınan toplam puanlar üzerinde aritmetik ortalama ve standart sapma hesaplamanın yanlış olacağını vurgulamıştır. Burke (1963) sıralama – eşit aralıklı ayrımını ölçme yönelimli (measurement-directed) ve ölçmeden bağımsız (measurement-independent) olarak sınıflamıştır. Ölçme yönelimli grup psikolojik testlerden elde edilen puanlarla yapılacak çıkarımsal istatistiklerde parametrik olmayan tekniklerin kullanılması gerektiğini savunurken, ölçmeden bağımsız grup normal dağılım gibi varsayımların karşılanması halinde toplam puanların eşit aralıklı kabul edilebileceğini savunmuştur. Lord ve Novick (1968) ise bu ikileme daha faydacı yaklaşmış, yerleştirme veya performansın yordanmasında verileri eşit aralıklı kabul etmek daha çok yarar sağlıyorsa toplam puanların eşit aralıklı kabul edilebileceğini belirtmiştir. Crocker ve Algina (2008) bu tartışmalara değindikten sonra Lord ve Novick'in (1968) görüşüne katıldıklarını ifade etmiştir. Günümüzde de testten alınan toplam puanlar normallik varsayımını karşılıyorsa genelde eşit aralıklı düzeyde kabul edilmektedir. Bu tartışmalar ışığında bölümün bundan sonraki kısımlarında testten alınan puanların eşit aralıklı ölçme düzeyinde olduğu kabul edilmiştir.

Bir dağılımı tanımlamak ve anlamlandırmak için iki temel ölçüye ihtiyaç duyulur. Bunlar dağılımdaki verilerin merkezinin ne olduğu ve bu verilerin merkezden ne kadar uzağa dağıldığıdır (Köklü ve ark., 2007). Sırasıyla merkezî eğilim ölçüleri ve yayılım ölçüleri olarak adlandırılan bu ölçüler bölümün devamında detaylı şekilde ele alınmıştır. Bahse konu ölçüleri birbiriyle karşılaştırabilir kılmak adına, Tablo 1'de örnek bir veri matrisine yer verilmiş ve tüm ölçüler bu matristen elde edilmiştir. Tabloda doğru yanıtlar 1; yanlış ve boş yanıtlar ise 0 olarak gösterilmiştir. Bu tür puanlamalar iki kategorili puanlama ya da 1 – 0 puanlama olarak adlandırılmaktadır.

Tablo 1

Beş Maddelik Bir Teste İlişkin Yanıt Matrisi

Öğrenci	Madde 1	Madde 2	Madde 3	Madde 4	Madde 5	Toplam
Ö1	1	1	1	1	1	5
Ö2	1	1	0	1	1	4
Ö3	1	1	1	1	0	4
Ö4	1	1	0	0	1	3
Ö5	1	1	1	0	0	3
Ö6	1	1	0	1	0	3
Ö7	1	1	0	0	1	3
Ö8	1	0	1	0	0	2
Ö9	1	0	0	0	1	2
Ö10	1	0	0	0	0	1

2.1. Merkezî Eğilim Ölçüleri

Dağılımın hangi noktaya eğilim gösterdiğini temsil eden ölçülerdir. Bu başlık altındaki ölçülerin her biri, merkezî eğilimi bulmayı amaçlasa da taşıdıkları anlamlar farklıdır. En sık kullanılan merkezî eğilim ölçüleri ortalama (eğitimdeki ölçmelerde genellikle aritmetik ortalama), ortanca ve tepe değerdir.

Dağılımda yer alan tüm değerlerin aritmetik ortalamayı etkilediği unutulmamalıdır. Başka bir deyişle aritmetik ortalama uç değerlerden etkilenir. Dolayısıyla dağılımda uç değerler varsa aritmetik ortalama yanlış yorumlamalara neden olabilecek bir merkezi eğilim ölçüsü haline gelir.

Aritmetik Ortalama ($\bar{X}$): Bir dağılımdaki tüm gözlemlerin toplanıp gözlem sayısına bölünmesiyle hesaplanır. *X* bir gözlemi, *n* ise gözlem sayısını temsil etmek üzere, aritmetik ortalamanın matematiksel gösterimi Eşitlik 1'de verilmiştir (Kaptan, 1998).

$$\bar{X} = \frac{\sum X}{n} \qquad \textit{Eşitlik 1}$$

Tablo 1'de yer alan örnek dağılımda, öğrencilerin puanlarına ait aritmetik ortalamayı elde etmek için tüm puanlar toplanıp gözlem sayısına bölünmelidir. Buna göre puanların aritmetik ortalaması:

$\bar{X} = \frac{5+4+4+3+3+3+3+2+2+1}{10} = \frac{30}{10} = 3$ olarak bulunur.

Ortanca: Medyan olarak da isimlendirilen bu istatistik bir dağılımdaki değerlerin tam orta noktasını gösterir. Yani ortanca, gözlemlerin sayısal değerlerinden değil sıralamasından etkilenir. Dolayısıyla bir gözlemin 5,50 ya da 100 olmasından ziyade, gözlemin kaçıncı sırada olduğu önemlidir. Bir dağılımın ortancası bulunurken dağılımdaki değerler küçükten büyüğe sıralanır ve ortadaki değer tespit edilir. Eğer dağılımda tek sayıda gözlem varsa ortada tek bir değer olacağından ortanca kolayca

belirlenir. Ancak dağılımda çift sayıda gözlem varsa ortada tek bir değer bulunmaz, bu durumda ortadaki iki değerin aritmetik ortalaması alınarak ortanca değere ulaşılır (Arıcı, 2006).

Tablo 1'deki toplam puanlar büyükten küçüğe doğru 5, 4, 4, 3, **3, 3**, 3, 2, 2, 1 şeklinde sıralanır. Bu dağılımın tam ortasında koyu ile yazılmış olan iki adet 3 gözlemi bulunur. Ortada iki adet gözlem olduğu için bu iki gözlemin aritmetik ortalaması dağılımın ortancasıdır. Dolayısıyla bu dağılımın ortancası (3+3) / 2 = 3'tür. Gözlem sayısının çok fazla olduğu durumlarda ortancayı bu örnekteki gibi kolayca bulmak mümkün değildir. Bunun yerine *n* gözlem sayısını temsil etmek üzere Eşitlik 2 kullanılarak ortancanın *kaçıncı gözlem* olduğu, diğer bir deyişle konumu (L) bulunabilir.

$$L = \frac{n+1}{2} \qquad \textit{Eşitlik 2}$$

Buna göre, örnekte yer alan 10 gözlem için ortancanın konumu (10 + 1) / 2 = 5,5 bulunur. Yani 5. ile 6. gözlemin ortası (aritmetik ortalaması) bu dağılımın ortancasıdır. Dağılımın 5. ve 6. gözlemleri sırasıyla 3 ve 3'tür. Bu iki değerin aritmetik ortalaması da 3 olduğundan dağılımın ortancası 3'e karşılık gelir.

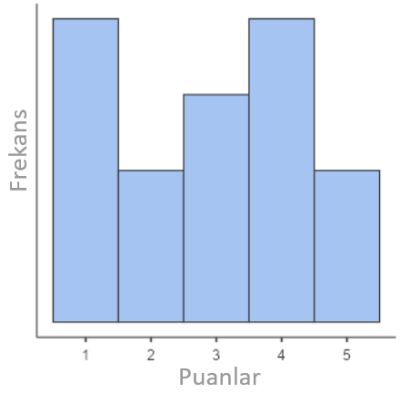


Ülkelerin gelirleri üzerine raporlama yapılırken genellikle aritmetik ortalama yerine ortanca tercih edilir. Bunun nedeni, ülke nüfusundaki az sayıda olan çok zenginlerin, gelir ortalamasını yukarı çekmesinin getireceği yanlış yorumlama riskidir. Gelirin çarpık dağılımı nedeniyle aritmetik ortalama yerine ortancanın raporlanması daha çok tercih edilir.

Tepe Değer: Tepe değer (mod), bir dağılımda en çok tekrar eden gözlemi temsil eder. Tablo 11.1'deki toplam puanlar incelendiğinde 1 adet 5; 2 adet 4; 4 adet 3; 2 adet 2 ve 1 adet 1 puanın olduğu görülür. Dolayısıyla bu seride en çok tekrar eden toplam puan değeri 3'tür ve dağılımın tepe değeri 3 olarak yorumlanır. Tepe değer de tıpkı ortanca gibi uç değerlerden etkilenmez. Bunun yerine değerlerin tekrar etme sıklığından etkilenir. Bir dağılımda tek bir tepe değer olabileceği gibi iki ya da daha fazla tepe değer de bulunabilir (Turgut & Baykul, 2010). Tepe değer, özellikle en sık tekrar eden yanıtın ya da puanın belirlenmesinde; dağılımın hangi noktalarda yoğunlaştığının incelenmesinde kullanışlılık sağlar. Örneğin; öğrenci puanlarının 1, 1, 1, 1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5 olduğu bir durumda, dağılımda en çok tekrar eden puanlar 4'er defa tekrarlanan 1 ve 4 olur. Bu puanlar için aritmetik ortalama hesaplandığında 43 / 15 = 2,87 değeri elde edilir. Öte yandan bu dağılımın ortancası 3'tür. Oysaki bu dağılımdaki veriler Şekil 1'deki gibi görselleştirildiğinde dağılımdan başka bir anlam çıkmaktadır.

Şekil 1

Örnek Dağılıma Ait Histogram



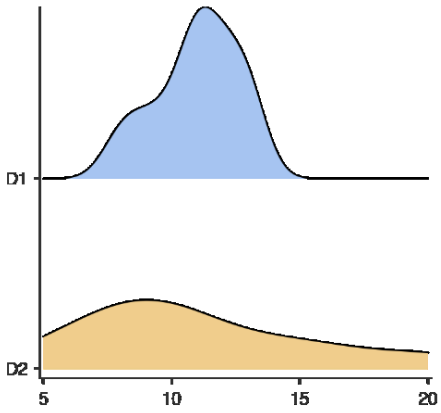
Şekil 1'de görüldüğü üzere dağılımın aritmetik ortalaması 2,87 ve ortancası 3 olmasına rağmen öğrenciler 1 ve 4 puanda yoğunlaşmıştır. Böyle dağılımlar bimodal (çift modlu / tepe değerli) olarak adlandırılmaktadır. Bu anlamda bir dağılımın tepe değeri, aritmetik ortalama ve ortancanın sağlayamadığı bilgileri de sunabilmektedir.

2.2. Yayılım Ölçüleri

Bir dağılımı yalnızca merkez noktası ile tanımlamak yeterli değildir. Dağılımdaki gözlemlerin merkez noktadan ne kadar uzağa dağıldığını da tespit etmek, dağılım hakkında daha fazla bilgi verir. Bu durum, dünya üzerindeki bir konumun yalnızca enlem ile değil aynı zamanda boylamla tanımlanmasına benzetilebilir. Örnek üzerinden açıklamak amacıyla Şekil 2'de aritmetik ortalaması aynı olan iki dağılıma ait yoğunluk grafiği verilmiştir.

Şekil 2

İki Dağılıma Ait Yoğunluk Grafiği



Şekil 2'deki iki dağılımın da aritmetik ortalaması 11'dir. Buna karşın üstte yer alan D1 dağılımındaki veriler 11 noktasına daha yakın dağılırken altta bulunan D2 dağılımındaki veriler 11 noktasından çok daha uzağa dağılmıştır. Yayılım ölçüleri, bir dağılımdaki gözlemlerin birbirinden ne kadar uzağa yayıldığı hakkında bilgi sağlamaktadır. Bu ölçüler aynı zamanda dağılım ölçüleri olarak da anılmaktadır. Dağılım ölçüleri birer homojenlik/heterojenlik ölçüsü olarak da nitelendirilmektedir.

Dağılım Açıklığı (Ranj): Dağılım açıklığı ya da ranj bir dağılımdaki en büyük ve en küçük değer arasındaki farka eşittir (Arıcı, 2006). Dağılıma ait kabaca ve hızlı bir fikir edinmek isteniyorsa kullanışlı bir ölçüdür. Dağılım açıklığı hesaplanırken en küçük ve en büyük değer kullanıldığı için verilerin hangi aralıkta değiştiğine dair test geliştiricisine fikir verir. Örneğin; öğrenci puanlarının 50 – 90 aralığında değiştiğini ve dağılım açıklığının 40 olduğunu bulan bir test geliştiricisi dağılımdaki en başarısız öğrencinin bile 50 puan aldığını, bunun yanında tüm maddelere doğru yanıt veren (dolayısıyla 100 puan alan) öğrenci bulunmadığını, en başarılı ile en başarısız öğrenci arasında ise 40 puanlık bir fark olduğunu söyleyebilir.

Varyans: Dağılımdaki gözlemlerin aritmetik ortalamadan ne kadar uzaklaştığına ilişkin bir ölçü olup gözlemlerin değişkenliği hakkında bilgi verir (Woolley, 1998). *X*, gözlenen değeri; $\bar{X}$ aritmetik ortalamayı ve *n* toplam gözlem sayısını temsil etmek üzere bir dağılımın varyansı Eşitlik 3 kullanılarak hesaplanır

$$s^2 = \frac{\sum(X - \bar{X})^2}{n - 1} \qquad \textit{Eşitlik 3}$$

Tablo 1'deki toplam puan dağılımına ait varyans hesaplanmak istendiğinde ve bu dağılımın aritmetik ortalamasının 3 olduğu hatırlanırsa;

$$s^2 = \frac{(5-3)^2 + (4-3)^2 + (4-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (2-3)^2 + (2-3)^2 + (1-3)^2}{10-1}$$

$$s^2 = \frac{(2)^2 + (1)^2 + (1)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2 + (-1)^2 + (-1)^2 + (-2)^2}{9}$$

$$s^2 = \frac{4 + 1 + 1 + 0 + 0 + 0 + 0 + 1 + 1 + 4}{9} = \frac{12}{9} = 1{,}33 \text{ olarak bulunur.}$$

Ancak bu hesaplamada toplam puanların birimi "puan" olarak kabul edilirse hesaplanan varyansın birimi "puan^2" olur. Tam da bu sebeple, varyans istatistiksel testlerde sıklıkla kullanılmasına karşın bir dağılımı tanımlayıp yorumlamak için pek işlevsel ölçü değildir.

Standart Sapma: Varyans biriminin, hesaplandığı gözlem biriminin karesi olması nedeniyle dağılıma dair yorumda bulunmak güçleşir. Bu nedenle birimi, toplanan verilerdeki gözlemlerin birimi ile aynı olacak bir ölçüye ihtiyaç duyulur. Standart sapma bu ihtiyaca cevap vermekte ve varyansın karekökü alınarak elde edilmektedir. Tablo 1'deki toplam puanlara ait varyansın 1,33 olduğu hatırlanırsa bu dağılıma ait standart sapma değeri $s = \sqrt{1{,}33} = 1{,}15$ olarak elde edilir.

Gerek varyans, gerekse standart sapma verilerin aritmetik ortalamadan sapma miktarını gösteren ölçüler olduğundan bu istatistiklere ait değerler büyüdükçe gözlemler ortalamadan uzaklaşır; küçüldükçe gözlemler ortalamaya yaklaşır. Bu bakımdan Şekil 2'de yer alan iki dağılımdan D1; D2'ye göre daha küçük bir standart sapmaya, dolayısıyla daha küçük bir varyansa sahiptir. Standart sapmanın küçülmesi, gruptaki öğrencilerin birbirine benzer ya da homojen olduğu şeklinde yorumlanırken standart sapmanın büyümesi gruptaki öğrencilerin birbirinden farklı ya da grubun heterojen olduğu şeklinde yorumlanır.

2.3. Standart Puanlar

Farklı ölçme sonuçlarını karşılaştırırken ham puanlar yerine ölçme sonuçlarının ortak bir birime dönüştürüldüğü standart puanlar kullanılır. Örneğin; matematik testinden 20, Türkçe testinden 15 doğrusu olan bir öğrencinin matematik testinde daha başarılı olduğunu söylemek yanıltıcı olabilir. Öyle ki bu öğrenci matematik testinde toplam 40 maddeden 20 doğru yaparken Türkçe testinde 20 maddeden 15 doğru yapmış olabilir. İşte, ham puanların standart puanlara dönüştürülmesi bu gibi farklı ölçme sonuçlarını karşılaştırılabilir kılar. Bunun için yaygın olarak kullanılan standart puan, z-puanıdır. Bu puan aynı zamanda öğrencilerin kendi içinde karşılaştırılmasına da olanak sağlar ve bir öğrencinin ortalamanın kaç standart sapma uzağında olduğu hakkında bilgi verir.

Standart z-Puanı: Ortalaması 0 ve standart sapması 1 olan bir dağılıma sahiptir ve bireyin ortalamanın kaç standart sapma uzağında olduğunu gösterir. Bir kişiye ait *z*-puanı; *X*, gözlenen puan; $\bar{X}$, aritmetik ortalama ve *S*, standart sapma olmak üzere Eşitlik 4'teki gibi hesaplanır (Gulliksen, 1950):

$$z = \frac{X - \bar{X}}{S} \qquad \textit{Eşitlik 4}$$

Tablo 1'de yer alan örnekteki toplam puanların aritmetik ortalamasının 3 ve standart sapmasının 1,15 olduğu hatırlanırsa her bir öğrencinin *z*-puanı aşağıdaki gibi hesaplanabilir:

Tablo 2

Öğrencilere Ait z-Puanları

Öğrenci	X	$X - \bar{X}$	$z = \frac{X - \bar{X}}{S}$
Ö1	5	2	1,74
Ö2	4	1	0,87
Ö3	4	1	0,87
Ö4	3	0	0,00
Ö5	3	0	0,00
Ö6	3	0	0,00
Ö7	3	0	0,00
Ö8	2	-1	-0,87
Ö9	2	-1	-0,87
Ö10	1	-2	-1,74

Yukarıda bahsedildiği gibi *z*-puanı, ortalaması 0 ve standart sapması 1 olan bir dağılımdır. Örnekte hesaplanan *z*-puanlarının da aritmetik ortalamasının 0 ve standart sapmasının 1 olduğu görülebilir. Standart *z*-puanına bakılarak bireyin ortalamanın kaç standart sapma uzağında olduğu şeklinde bir yorum yapılabilir. Örnekteki 1. öğrenci (Ö1), aritmetik ortalamanın 1,74 standart sapma üzerindeyken 9. öğrenci (Ö9) aritmetik ortalamanın 0,87 standart sapma altında bir performans sergilemiştir. Ö4-Ö7 arasındaki öğrencilerin puanı ise tam olarak aritmetik ortalamada olduğu için *z*-puanları sıfır olarak bulunmuştur.

Yukarıda açıklanan yöntem ham puanları doğrusal bir şekilde *z*-puanına dönüştürmektedir. Bu nedenle, hesaplanan *z*-puanları ham puanların sahip olduğu dağılımı korur. Öte yandan ham puanlar standart normal dağılım ile elde edilen yüzdelikler üzerinden doğrusal olmayan bir şekilde normalleştirilerek de *z*-puanına dönüştürülebilir. Bu durumda dönüşüm yapılan *z*-puanı dağılımı, normale yakın bir dağılım sergiler. Sözü edilen dönüşümü yapmak için öncelikle ham puanlardan yüzdelik dilimler oluşturulur, ardından her bir ham puanın yüzdelik dilimi standart normal dağılım altında kalan alan olarak ele alınır ve bu alana karşılık gelen *z*-puanı belirlenir (Crocker & Algina, 2008).

Standart T-Puanı: Standart *z*-puanı teorik olarak -∞ ile +∞ arasında değişmekle birlikte, genelde -3 ile +3 arasında kalan değerler alır. Bu değerler test geliştiricisi tarafından kolayca yorumlanabilse de testi alanlar açısından yorumlaması güçtür. Bundan dolayı testi alanlar tarafından da kolay bir şekilde yorumlanabilen, ortalaması 50 ve standart sapması 10 olan standart T-puanı ileri sürülmüştür. T-puanı, Eşitlik 5'teki gibi *z*-puanı 10 ile çarpılıp sonuca 50 eklenerek elde edilmektedir (Gulliksen, 1950).

$$T = 50 + 10.z \qquad \textit{Eşitlik 5}$$

Buna göre, *z*-puanının hesaplandığı örnekteki her bir öğrenci için T-puanı hesaplanırsa Tablo 3'teki bulgular elde edilir:

Tablo 3

Öğrencilere Ait z ve T-Puanları

Öğrenci	X	X - $\bar{X}$	$Z = \frac{X - \bar{X}}{S}$	T = 50 + 10.Z
Ö1	5	2	1,74	67,4
Ö2	4	1	0,87	58,7
Ö3	4	1	0,87	58,7
Ö4	3	0	0,00	50,0
Ö5	3	0	0,00	50,0
Ö6	3	0	0,00	50,0
Ö7	3	0	0,00	50,0
Ö8	2	-1	-0,87	41,3
Ö9	2	-1	-0,87	41,3
Ö10	1	-2	-1,74	32,6

Görüldüğü üzere, hesaplanan T-puanları testi alanlar için daha alışıldık, dolayısıyla daha kolay yorumlanabilir hale gelmiştir.

3. MADDE PUANLARI ÜZERİNDE YAPILAN İSTATİSTİKSEL İŞLEMLER

Test puanları, testi oluşturan madde puanları ile yakından ilişkilidir. Test puanları üzerinde yapılan istatistiksel işlemler tüm gruba (örneğin, bir sınıfın başarı düzeyi) ait bilgi sunarken madde puanları üzerinde yürütülen istatistiksel işlemler, maddelerin uygulamanın yapıldığı grupta nasıl çalıştığı hakkında bilgi verir. Madde güçlüğü ile madde ayırt ediciliği sıklıkla kullanılan iki madde istatistiği olup madde standart sapması, madde varyansı, madde güvenirliği gibi istatistikler de bulunmaktadır. Sıklıkla kullanılan madde istatistikleri, kitabın 2. bölümünde bahsedilen iki ana kurama göre aşağıda ayrı ayrı detaylı olarak ele alınmıştır.

ÖSYM'nin yaptığı birçok testin raporlanmasında standart puanlar kullanılmaktadır. Bu testlerden biri olan Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı'nın (ALES) Başvuru Kılavuzu incelendiğinde adayların ALES puanının aşağıdaki eşitlikle hesaplandığı görülmektedir (ÖSYM, 2025):

$$ALES\ puanı = 70 + \frac{30[2(AP - X) - S]}{2(B - X) - S}$$

AP: Adayın ağırlıklı puanı
X: Ağırlıklı puanların ortalaması
S: Ağırlıklı puanların standart sapması
B: Ağırlıklı puanların en büyüğü

Aynı kılavuzda adayın ağırlıklı puanının hesaplanması hakkında, "Adayların her testteki sorulara verdikleri doğru ve yanlış cevaplar ayrı ayrı toplanacak, doğru cevap sayısından yanlış cevap sayısının dörtte biri çıkarılarak ham puanlar elde edilecektir. Her test için adayların ham puanlarının ortalaması ve standart sapması ayrı ayrı hesaplanarak adayların ham puanları, ortalaması 50 ve standart sapması 10 olan standart puanlara dönüştürülecektir." ifadesi yer almaktadır. Dolayısıyla, adayların ham puanları öncelikle T-puanına (ortalaması 50, standart sapması 10) dönüştürülmektedir. Bunu takiben dönüştürülen bu T-puanları her bir puan türüne göre ağırlıklandırılarak adayın ağırlıklı puanı elde edilmektedir. Ardından tüm adayların ağırlıklı puan ortalaması, standart sapması ve en yüksek ağırlıklı puan kullanılarak ALES puanı raporlanmaktadır. Okuyucular dikkat edecek olursa bu sürecin hiçbir noktasında maddenin ne kadar zor olduğu hesaplamaya dâhil edilmemektedir. Ancak halk arasında "zor sorunun daha çok puanlandığına" yönelik bir yanılgı vardır. Kitabın 2. bölümünde bu yanılgıya ait bir tartışmaya değinilmiştir.

3.1. Madde Güçlüğü

Bir testteki maddelerin, testi alan gruba ne kadar zor geldiğini belirlemek önemlidir. Maddeyi yanıtlayan grubun maddenin ölçtüğü özelliğe ne derece sahip olduğunu ortaya koymanın yanı sıra, testin geliştirilme amacına uygun olan maddeleri tespit edebilmek için de madde güçlüğünü hesaplamak gerekir. Aşağıda hem KTK temelinde madde güçlük indeksi, hem de MTK temelinde madde güçlük parametresine yer verilmiştir.

Klasik Test Kuramına Göre

Bir maddenin, uygulandığı gruba ne kadar kolay ya da zor geldiğini belirlemek için maddenin güçlüğü incelenir. Madde güçlüğü, maddeyi doğru yanıtlayan birey sayısının, maddeyi yanıtlayan tüm bireylere oranını ifade eder. Bu oran 1-0 şeklinde puanlanan maddeler için aynı zamanda madde puanlarının ortalamasına eşittir (Erkuş, 2012). Maddeyi grubun tamamı doğru yanıtlarsa madde güçlüğü 1değerini alır, diğer taraftan gruptan hiç kimse maddeye doğru yanıtlamazsa maddenin güçlüğü 0 olur. Dolayısıyla bu indeksin adı maddenin *güçlüğü* olsa da indeks değeri arttıkça madde kolaylaşır. İndeks değeri 0,50 olduğunda ise madde tam orta güçlüktedir. Madde güçlük indeksi p_j sembolü ile gösterilir ve j madde numarasını temsil etmek üzere Eşitlik 6 aracılığıyla hesaplanabilir:

$$p_j = \frac{\overline{X}_j - min_j}{max_j - min_j} \qquad \textit{Eşitlik 6}$$

Bu eşitlikte $\overline{X}_j$, güçlüğü hesaplanmak istenen maddenin ortalama puanını; min_j o maddeden alınabilecek en düşük puanı ve max_j ise o maddeden alınabilecek en yüksek puanı temsil eder. Buna göre, 1-0 şeklinde puanlanan bir madde için $max_j = 1$ ve $min_j = 0$ olacağından Eşitlik 6 otomatik olarak $\overline{X}_j$'ye dönüşür. $\overline{X}_j$ hesaplanırken her bir puan (ya 0 ya da 1) toplanıp toplam kişi sayısına bölündüğünden 1-0 puanlanan maddeler için güçlük, "*doğru yanıtlayan kişi sayısı / toplam kişi sayısı*" formülüyle hesaplanır. 1-0 dışında puanlanan maddeler (örneğin, açık uçlu maddeler) içinse o maddeden alınan puan ortalamalarından, maddeden alınabilecek en düşük puan (genellikle 0) çıkarılıp elde edilen fark maddenin beklenen ranjına bölünür. Örneğin, alınabilecek en yüksek puanın 10, en düşük puanın 0 olduğu bir açık uçlu maddeden alınan puanların aritmetik ortalaması 6,70 ise bu maddenin güçlüğü: (6,70 – 0) / (10 – 0) = 0,67 olarak bulunur.

Yukarıda aktarılan genel eşitlik (eşitlik 6) haricinde, en başarılı ve en başarısız %27'lik katılımcı grubunu dikkate alarak da madde güçlüğünü hesaplamak mümkündür. Alt ve üst %27'lik gruplar yöntemine göre madde güçlüğü Eşitlik 7 ile hesaplanır.

$$p_j = \frac{n_{d\ddot{u}} + n_{da}}{n_{\ddot{u}} + n_a} \qquad \textit{Eşitlik 7}$$

$n_{d\ddot{u}}$*:* Üst %27'lik grupta maddeye doğru yanıt veren kişi sayısı
n_{da}*:* Alt %27'lik grupta maddeye doğru yanıt veren kişi sayısı
$n_{\ddot{u}}$ *ve* n_a*:* Üst ve alt %27'lik gruplardaki kişi sayısı.



Alt ve üst gruptaki kişileri belirlerken mutlaka %27'lik dilimi referans almak gibi bir zorunluluk yoktur. Madde analizi için en sık başvurulan dilim %27 olmakla birlikte; Murphy ve Davidshofer (2014) alt ve üst grubun testin uygulandığı örneklemin %25 ile %35'ini kapsayabileceğini belirtmiş, bu aralıktaki farklı yüzdelik dilimlerin benzer sonuçlar üreteceğini vurgulamıştır.

Test geliştirme sürecinde madde güçlüğünün nasıl olması gerektiği testin amacına bağlıdır. Bu noktada madde varyansı, yani maddenin testi alan bireyler arasındaki farklılığı ne derecede ortaya koyabildiği ön plana çıkmaktadır. Madde varyansı bir maddenin güçlüğü ile ilişkili olup "pj × qj" ile hesaplanır. Bu işlemdeki *qj*, 1 - pj'dir. Yani bir maddenin varyansı pj(1 – pj) eşitliği ile bulunur. Bu bağlamda, madde varyansı madde güçlüğünün 0,50 olduğu durumda en yüksek değerini alır ve 0,50(1 - 0,50) = 0,25 olur. Bunun dışındaki madde güçlüklerinde madde varyansı 0,25'ten daha küçük çıkar. Örneğin, madde güçlüğünün 0,10 olduğu durumda madde varyansı 0,10(1 – 0,10) = 0,09 olarak elde edilir. Madde güçlüğünün 0,10 olması; bu maddenin testi alan grubun %10'u tarafından doğru, %90'ı tarafından ise yanlış yanıtladığını gösterir. Bahse konu madde testi alanlar arasındaki farklılığı ortaya koymakta, güçlüğü 0,50 olan bir maddeye kıyasla yetersiz kalmaktadır.

Öte yandan özellikle çoktan seçmeli maddelerde şans başarısı söz konusudur. Örneğin, beş seçenekli bir maddenin şansla doğru yanıtlanma olasılığı 1 / 5 = 0,20'dir. Bu durum göz önüne alındığında bireyler arasındaki farklılığı en çok ortaya çıkarması beklenen madde güçlük düzeyinde bir düzeltme gerekir. Bunun için, *m* seçenek sayısını göstermek üzere madde güçlüğüne 0,50 / m eklenmelidir (Crocker & Algina, 2008). Buna göre beş seçenekli bir madde için bireyler arasındaki farklılığı en fazla ortaya koyması beklenen madde güçlüğü 0,50 değil, 0,50 + 0,50 / 5 = 0,60'tır. Lord (1952) ise yaptığı simülasyon çalışmasında hesaplanan bu değerden biraz daha kolay olan maddelerin seçilmesi halinde en yüksek test güvenirliğine ulaşıldığını ifade etmiştir.

Yukarıdaki hesaplama, madde varyansını en yüksek yapması beklenen madde güçlüğüne yönelikti. Benzer şekilde şans başarısı dikkate alınarak hesaplanan madde güçlüğünde bir düzeltme yapılmak istenirse Eşitlik 8 kullanılabilir (Erkuş, 2003):

$$p_j' = \frac{p_j - c}{1 - c}$$

p_j': *Düzeltilmiş madde güçlüğü*
c: *Maddenin şansla doğru yanıtlanma olasılığı*

Eşitlik 8

Buna göre, beş seçenekli bir madde için madde güçlüğü 0,40 olarak elde edildiyse düzeltilmiş madde güçlüğü: p_j' = (0,40 – 0,20) / (1-20) = 0,25 bulunur. Dikkat edilirse düzeltilmiş madde güçlük indeksi (0,25), ilk hesaplanan madde güçlüğü indeksine (0,40) göre azalmış, daha açık bir anlatımla maddenin gerçekte daha zor olduğu belirlenmiştir. Bu eşitlik az önce beş seçenekli bir madde için varyansı en yüksek değere ulaştırması beklenen 0,60 değerine uygulanırsa (0,60 – 0,20) / (1 – 0,20) = 0,50 değeri elde edilir. Öte yandan Lord (1952) maddenin şansla doğru yanıtlanma olasılığına dayalı madde güçlüğünün hesaplanmasında, özellikle düzeltme formülü kullanıldığı durumda (adayın doğru sayısından yanlış sayısının seçenek sayısının bir eksiğine bölünerek çıkarılması halinde), adayların tahminde bulunma davranışının azaldığına; bu nedenle hesaplamada tam olarak 1/m kullanmanın uygulamadaki sınırlılığına dikkat çekmiştir.

Testteki maddelerin güçlük indekslerinin toplamı ise testin ortalamasını vermektedir. Dolayısıyla bireyler arasındaki farklılığı en üst düzeyde ortaya koymak için ya tüm maddeler orta güçlükte olmalı ya da madde güçlük indeksleri farklı değerler alsa da (zor ya da kolay maddeler) ortalama test güçlüğünün 0,50 civarında olması sağlanmalıdır (Erkuş, 2003). Maddeler, güçlük indekslerine göre kolay-orta-zor şeklinde bir sınıflandırmaya tabi tutulmak istendiğinde, herkes tarafından kabul gören sınırlar bulunmamaktadır. Popham (2014), 0,20'den küçük değerlerin zor; 0,80 ve üzeri değerlerin kolay madde olarak sınıflandırıldığını ifade etmiş ama bu sınıflandırmaya karşı nötr bir tavır takınmıştır. Çünkü bir maddeye ait güçlüğün 0,90 olarak bulunması, maddenin uygulandığı grubun %90'ı tarafından doğru yanıtlandığını, dolayısıyla gruba kolay geldiğini bildirmesinin yanında grubun %90'ının maddenin ölçtüğü kazanıma ulaştığını da gösterir. Bu bakımdan yalnızca indeks değerini yorumlamak yerine, maddenin hangi bağlamda uygulandığına göre bir inceleme yapmak gerektiğini ifade etmiştir.

Alanyazında madde güçlüğünü sınıflandırmada kullanılabilecek keskin eşik değerlerden kaçınılsa da uygulayıcıların farklı kesme noktaları tanımladığı görülmektedir. Örneğin, Assessment System Corporation; 0,00 – 0,30 aralığını çok zor, 0,31 – 0,50 arasını zor, 0,51 – 0,70 arasını orta, 0,71 – 0,90 arasını kolay ve 0,91 – 1,00 arasını çok kolay madde olarak yorumlamaktadır (Thompson, 2020). Bu anlamda madde güçlük indeksi için 0,00 – 0,20 çok zor, 0,21 – 0,40 zor, 0,41 – 0,60 orta, 0,61 – 0,80 kolay ve 0,81 – 1,00 çok kolay kategorizasyonu pratikte kullanılabilecek bir sınıflandırmadır.

Madde Tepki Kuramına Göre

Örtük özellikler kuramı olarak da adlandırılan Madde Tepki Kuramı (MTK), bireyin bir maddeye vereceği yanıtın olasılığı ile yetenek düzeyi arasındaki ilişkiyi genellikle lojistik bir fonksiyonla inceler (Hambleton & Swaminathan, 1985). MTK'nin KTK ile benzerlikleri, KTK'ye olan üstünlük ve sınırlılıkları kitabın 2. bölümünde detaylı biçimde verilmiştir. Bu bölümde yalnızca MTK'ye dayalı madde güçlüğü ile ayırt ediciliğine ilişkin bazı hatırlatmalar yapılıp hesaplama ve yorumlama üzerinde durulmuştur.

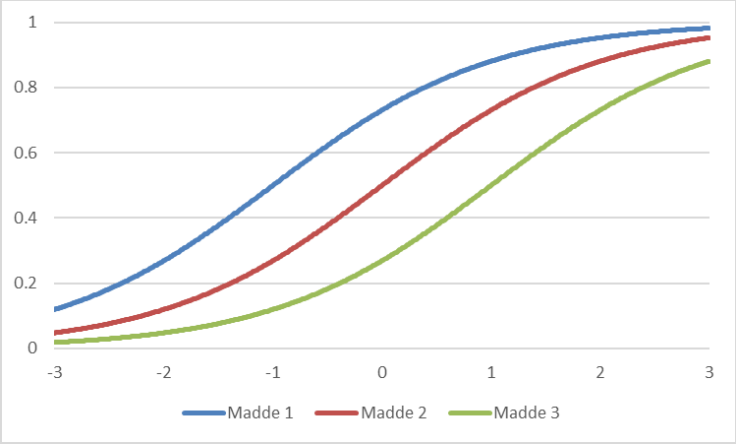
MTK'ye göre madde güçlüğü, maddeyi doğru yanıtlama olasılığının %50 olduğu yetenek düzeyini temsil eder ve *b parametresi* olarak adlandırılır (Baker & Kim, 2017). Örneğin, bir maddenin *b parametresi* -1 ise yetenek düzeyi -1 olan bir bireyin bu maddeyi doğru yanıtlama olasılığı %50'dir. Anlaşılacağı üzere MTK'de madde parametresi, yetenek düzeyi ile temsil edilir. Başka bir deyişle madde ve birey özellikleri aynı ölçekte yer alır.

MTK'ye göre madde güçlük parametresinin kestirimi, KTK'de yapılan madde güçlük indeksi hesaplanmalarından çok daha karmaşıktır. Bölümün sadeliğini koruyabilmek adına aşağıda *b parametresi*nin kestirimi yerine, bu parametrenin yorumu üzerine durulmuştur. Ancak ilgili okuyucular Marjinal En Yüksek Olabilirlik (Marginal Maximum Likelihood), Birleşik En Yüksek Olabilirlik (Joint Maximum Likelihood) gibi yöntemleri araştırabilir.

MTK'de *b parametresi*nin yetenek düzeyi ile aynı ölçekte ifade edilmesi nedeniyle *b parametresi*nin değeri azaldıkça madde kolaylaşır (maddeyi %50 olasılıkla doğru yanıtlamak için daha az yetenek gerekir), *b parametresi*nin değeri arttıkça madde zorlaşır. Buna bağlı olarak KTK ve MTK'deki madde güçlüğü değerleri ters şekilde yorumlanır. *b parametresi* yetenek düzeyi ile aynı ölçekte ele alındığından teorik olarak -∞ ile +∞ arasında uzanabilir. Ancak pratikte genellikle -3 ile +3 ya da -4 ile +4 arasında değerler alır. Şekil 3'te *b parametresi* sırasıyla -1, 0 ve 1 olan üç maddeye ait madde karakteristik eğrileri verilmiştir.

Şekil 3

Üç Maddeye Ait Madde Karakteristik Eğrisi



Madde karakteristik eğrileri incelenirse Madde 1'e ait eğri daha solda; madde 3'e ait eğri ise en sağdadır. Bunun nedeni, eğrilerin orta noktalarının madde 1 için -1.00 ve madde 3 için 1.00'da yer almasıdır. Buna göre en kolay madde 1. madde iken (çünkü bu maddeyi doğru yanıtlamak için gerekli yetenek düzeyi daha düşüktür), 3. madde en zor maddedir (bu maddeyi doğru yanıtlamak için gerekli yetenek düzeyi daha yüksektir).



Bu noktalar aynı zamanda madde karakteristik eğrilerinin dönüş yaptığı noktalara denk gelmektedir. *b parametresi*nin olduğu noktada madde karakteristik eğrisinin yönünü değiştirdiğine dikkat edilmelidir.

3.2. Madde Ayırt Ediciliği

Bir testte yer alan maddelerin ölçülen özelliğe az ya da çok sahip olanları ayırabilmesi istenir. Örneğin, testin genelinde başarılı olan öğrenciler maddeyi doğru yanıtlarken testin genelinde başarısız olan öğrencilerin maddeyi yanlış yanıtlaması beklenir. Bir maddenin bunu ne derece başardığını belirlemek için madde ayırt edicilik indeksi kullanılır. Aşağıda hem KTK temelinde madde ayırt edicilik indeksine hem de MTK temelinde madde ayırt edicilik parametresine yer verilmiştir.

Klasik Test Kuramına Göre

KTK'ye göre madde ayırt edicilik indeksi r_{jx} sembolü ile gösterilir. Bu sembolde *r* korelasyonu; *j* maddeyi ve *x* toplam puanı temsil eder. Dolayısıyla madde ayırt edicilik indeksi bir maddeden alınan puanlarla, testten alınan toplam puanlar arasındaki korelasyon katsayısı olarak ifade edilir. Madde ayırt edicilik indeksinin hesaplanmasında kullanılacak korelasyon katsayısı, madde türüne ve test geliştiricisinin altta yatan varsayımlarına bağlıdır. İkili puanlanan (1 – 0) maddeler için madde puanları iki kategorili, toplam puan ise süreklidir. Haliyle iki kategorili süreksiz bir değişken ile sürekli bir değişken arasındaki korelasyon katsayısının hesaplanması gerekir. Bu amaçla çift serili (r_{bis}) ya da nokta çift serili (r_{pbis}) korelasyon katsayısından yararlanılır. Madde puanlarının (1 – 0) gerçek süreksiz olduğu varsayılıyorsa nokta çift serili; yapay süreksiz (maddenin ölçtüğü özellik miktarı aslında sürekli, ancak yapay olarak 1 ve 0 olduğu) olduğu varsayılıyorsa çift serili korelasyon katsayısı kullanılır (Crocker & Algina, 2008). Maddeler çok kategorili puanlanıyorsa (örneğin, 0 – 10 arası puan), doğrudan Pearson momentler çarpımı korelasyon katsayısı ya da Spearman sıra farkları korelasyon katsayısı kullanılabilir. Ayrıca korelasyonda "madde – toplam" yerine düzeltilmiş madde toplam korelasyonunun kullanılması önerilir. Düzeltilmiş madde toplam korelasyonu, "madde – kalan" korelasyonu olarak da isimlendirilir ve herhangi bir maddeden alınan puanlar ile o madde haricinde testin kalan maddelerinden alınan toplam puan arasındaki korelasyonu gösterir. Madde – toplam korelasyonu yerine madde – kalan korelasyonun tercih edilmesinin nedeni, toplam puana ayırt ediciliği incelenen maddeye ait puanların da dâhil edilmesinin hesaplanacak madde korelasyonunu yapay olarak yükseltmesidir (Metsamuuronen, 2020).

Bunun yanında gruplar arası farkı (ya da oranları) temel alan alt – üst %27'lik gruplar yöntemi, Brennan indeksi ve Brennan'ın özel durum indeksi gibi madde ayırt edicilik indeksleri de vardır. Bu yöntemler korelasyon katsayısına dayalı olmadığı için r_{jx} yerine D veya B gibi harflerle gösterilir. Alt – üst %27'lik gruplar yöntemi, hesaplama kolaylığı nedeniyle özellikle elle hesaplamada tercih edilen bir yöntemdir ve Eşitlik 9 aracılığıyla hesaplanır (Popham, 2014).

$$D = p_{ü} - p_{a} \qquad \textit{Eşitlik 9}$$

Bu eşitlikte $p_{ü}$, üst grubun maddeyi doğru yanıtlama oranını; p_a ise alt grubun maddeyi doğru yanıtlama oranını temsil eder. Bu eşitliği kullanmak için öncelikle testi alan bireyler toplam puanlarına göre büyükten küçüğe doğru sıralanır. Ardından testi alan toplam kişi sayısının %27'si belirlenir ve puanı en yüksek olan %27'lik grup ile en düşük olan %27'lik grup seçilir. Daha sonra her bir madde için üst grupta maddenin doğru yanıtlanma oranından, alt grupta maddenin doğru yanıtlanma oranı çıkarılır. Bu yöntemde kullanılan %27 oranı ise Kelley (1939) tarafından gruptaki bireyler arası farkı en fazla ortaya çıkaran oran olarak bulunmuş ve ardından diğer araştırmacılar tarafından kullanılmaya başlanmıştır.

Brennan'ın indeksinde ise amaç bir maddenin mutlak bir ölçütün altında ve üstünde yer alan bireyleri ne kadar iyi ayırt ettiğini belirlemektir. Bu yönüyle daha önce aktarılan katsayılardan ayrılmaktadır. Şimdiye kadar yer alan madde ayırt edicilik indeksleri bir maddenin grubun genelini nasıl ayırdığına odaklanırken Brennan indeksi grubun belli bir kesiminin diğerlerinden nasıl ayrıldığına odaklanır ve Eşitlik 10 ile hesaplanır (Brennan, 1972).

$$B = \frac{U}{n_1} - \frac{L}{n_2}$$

n_1: Ölçüt puanın üzerindeki birey sayısı
n_2: Ölçüt puanın altındaki birey sayısı
U: Ölçüt puanın üzerinde puan alan bireylerden maddeye doğru yanıt veren sayısı
L: Ölçüt puanın altında puan alan bireylerden maddeye doğru yanıt veren sayısı

Eşitlik 10

Brennan indeksi sayesinde bir maddenin belli bir yeterlik düzeyinin üzerindeki ve altındaki bireyleri ne derece ayırt edebildiği saptanabilir. Arıkan ve Aybek (2022) maddenin belli bir yeterlik düzeyi yerine testi alan grubun belli bir bölümünü birbirinden ne derece ayırabildiğini belirlemek için Brennan'ın indeksini özel koşullar için uyarlamış ve önerdikleri uyarlamanın özellikle az sayıda kişinin seçilmesinin amaçlandığı durumlarda diğer indekslere göre daha etkili olduğunu ortaya koymuşlardır. Buna göre özel durumlar için Brennan'ın indeksi Eşitlik 11'deki gibi hesaplanmaktadır.

$$B_{x-y} = p_x - p_y$$

p_x: Üst %x grup için madde güçlüğü ya da maddeye doğru yanıt verme oranı
p_y: Kalan %y grup için madde güçlüğü ya da maddeye doğru yanıt verme oranı

Eşitlik 11

Bu eşitliğe göre, örneğin grubun en başarılı %10'luk grubunu kalan %90'lık grubundan ne derece ayırdığını belirlemek için:

$$B_{10-90} = p_{10} - p_{90}$$

eşitliği kullanılır. En başarılı %10'luk grubun maddeyi doğru yanıtlama oranından, kalan grubun maddeyi doğru yanıtlama oranı çıkarılır. Özellikle az sayıda öğrencinin seçildiği testler için B_{x-y}, belli bir grubun diğerlerinden ayırt edilmesi bakımından diğer madde ayırt edicilik indekslerine göre daha doğru sonuçlar vermektedir (Arıkan & Aybek, 2022).

Madde ayırt edicilik indeksi bir korelasyon katsayısı olduğundan -1 ile +1 arasında değerler alabilir. Madde ayırt edicilik indeksinin negatif olması, maddenin ters çalıştığını gösterir. Başka bir deyişle başarısı düşük olanlar maddeyi doğru yanıtlarken başarısı yüksek olanlar maddeyi yanlış yanıtlamıştır. Bu durumda akla ilk gelmesi gereken, cevap anahtarının yanlış hazırlanmış olması ihtimalidir. Cevap anahtarında herhangi bir problem yoksa maddenin testi alan öğrenciler tarafından anlaşılmamış ya da yanlış anlaşılmış olma olasılığı üzerinde durulmalıdır.

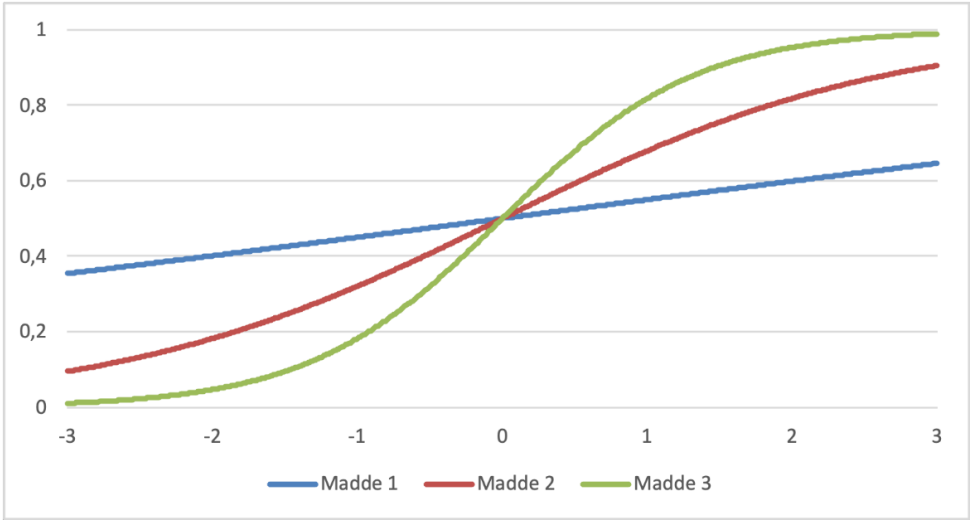
Test geliştirme sürecinde madde ayırt edicilik indeksinin 0,40'tan büyük olması beklenir. Madde ayırt edicilik indeksi 0,30 ile 0,39 arasında ise maddede düzeltme gerekmez ya da çok az düzelme gerekir. Madde ayırt edicilik indeksinin 0,20 ile 0,29 arasında olması halinde maddede düzeltme gerekirken 0,19 ve altındaki indekse sahip maddelerin testten çıkarılması gerekir (Popham, 2014). Bunun yanı sıra bir madde tüm katılımcılar tarafından doğru yanıtlandığında madde ayırt edicilik indeksi 0 olur. Bu durumda sıralanan ölçütler maddenin testten çıkarılması gerektiğine işaret eder. Fakat özellikle sınıf içi uygulamalarda bir maddenin bütün öğrenciler tarafından doğru yanıtlanması, tüm öğrencilerin maddenin ölçtüğü kazanıma ulaştığı anlamına gelir. Dolayısıyla yukarıda bahsedilen ölçütler testin kullanım amacı bağlamında değerlendirilmelidir.

Madde Tepki Kuramına Göre

MTK'de madde ayırt ediciliği için *a parametresi* kullanılır. Tek boyutlu MTK modellerinde iki kategorili ya da çok kategorili puanlanan maddeler için tek bir *a parametresi* bulunur. Bu parametrenin değeri büyüdükçe maddenin ayırt ediciliği artar, ancak bu değer çok yüksek olduğunda madde oldukça dar bir yetenek aralığı için ayırt edici olur ve bu aralık dışında kalan bireylerin büyük bölümü bu maddeyi ya doğru ya da yanlış yanıtlar (Baker, 2001). Şekil 4'te *b parametresi* 0 olan; *a parametresi* ise sırasıyla 0,20, 0,75 ve 1,50 olan üç maddeye ait madde karakteristik eğrileri verilmiştir.

Şekil 4

Ayırt Edicilikleri Farklı Üç Maddeye Ait Madde Karakteristik Eğrileri



Üç maddeye ait karakteristik eğrileri incelendiğinde Madde 1'e ait eğrinin oldukça yatay seyrettiği görülmektedir. Dolayısıyla bu madde, farklı yetenek düzeyindeki bireyleri ayırt etmede oldukça yetersizdir. Öyle ki yetenek düzeyi -2 olan (evrenin %2'sinden daha başarılı) bir bireyin bu maddeyi doğru yanıtlama olasılığı yaklaşık %40 iken yetenek düzeyi +2 olan (evrenin %97'sinden daha başarılı) bir bireyin bu maddeyi doğru

yanıtlama olasılığı yaklaşık %60'tır. Başka bir deyişle yetenek düzeyinin farklılaşması bu maddeyi doğru yanıtlama olasılığında önemli bir değişikliğe yol açmamaktadır. Öte yandan madde 3 incelendiğinde yetenek düzeyi -2 olan bir bireyin bu maddeyi doğru yanıtlama olasılığının yaklaşık %10; yetenek düzeyi +2 olan bir bireyin ise bu maddeyi doğru yanıtlama olasılığının yaklaşık %90 olduğu anlaşılmaktadır. Buna göre, bireylerin yetenek düzeyleri bu örnekteki 3. maddenin doğru yanıtlanma olasılığını önemli düzeyde etkilemektedir.

Bir maddenin *a parametresi* için Baker (2001) tarafından eşik değerler önerilse de pratikte *a parametresi*nin 1'den büyük olması istenir. Özellikle bir maddenin, birey hakkında ne düzeyde bilgi verdiğini temsil eden madde bilgi fonksiyonu hesaplanırken *a parametresi*nin karesi çarpan olarak hesaplamaya dâhil edilir. Bu nedenle *a parametresi*nin 1'den küçük olması halinde, hesaplanan bilgi fonksiyonu oldukça küçük değerler almaya başlar. Aşağıda üç farklı *a parametresi*nin madde bilgi fonksiyonuna katkıları hesaplanmıştır:

Tablo 4

a Parametresi ve Madde Bilgi Fonksiyonuna Katkı Miktarları

	a = 0,85	*a = 1,00*	*a = 1,15*
Madde bilgi fonksiyonuna katkı (a^2)	0,723	1,000	1,323

Görüldüğü üzere, *a parametresi* 1'den küçük değerler aldığında madde bilgi fonksiyonuna katkısı kendi değerinden daha az olurken 1'den büyük değerler aldığında madde bilgi fonksiyonuna katkısı kendi değerinden daha büyük olmaktadır. Öte yandan çok yüksek değere sahip bir *a parametresi* maddenin oldukça dar bir aralığı ayırt ettiği şeklinde yorumlanır.

Madde Tepki Kuramına Göre Diğer Madde Parametreleri

Detayları kitabın 2. bölümünde verilen ikili puanlanan maddeler için kullanılan MTK modellerinde *a* ve *b* parametrelerinin yanında *c* ve *d* parametresi de kullanılmaktadır. Bilindiği üzere *c* parametresi, maddenin ölçtüğü özelliğe sahip olmayan bireyin o maddeyi şansla doğru yanıtlamasına ilişkin bilgi verirken *d* parametresi maddenin ölçtüğü özelliğe sahip olan bireyin dikkatsizlik gibi nedenlerle maddeyi yanlış yanıtlama olasılığına ilişkin bilgi sağlar.

Kullanılan MTK modeli, *a, b, c* ve *d* parametrelerinden kaç tanesinin değişken olarak ele alındığına göre adlandırılır. Yalnızca *b* parametresi değişken olarak alınıyorsa 1 Parametreli Lojistik Model (PLM), *b* ve *a* parametreleri değişken olarak alınıyorsa 2PLM; *b, a* ve *c* parametreleri değişken olarak alınıyorsa 3PLM; *b, a, c* ve *d* parametreleri değişken olarak alınıyorsa 4PLM adını almaktadır. Pratikte 1, 2 ve 3PLM kullanılsa da 4PLM daha çok kuramsal çalışmalarda kullanılmaktadır.

3.3. Yanıtların Seçeneklere Dağılımı (Çeldiricilerin Analizi)

Maddelere ait güçlük ve ayırt edicilik indeksleri ya da parametreleri maddelerin nasıl çalıştığı hakkında fikir verir. Hesaplanan indeksler ya da parametreler doğrultusunda bazı maddelerde düzeltmelere gidilebilir. Bu noktada yalnızca madde kökü değil seçeneklerin, özellikle de çeldiricilerin incelenmesi, madde hakkında daha derin fikir sahibi olunmasını sağlar. Daha açık bir anlatımla, çoktan seçmeli bir maddede çeldiriciler yeterince iyi değilse başarısız öğrenci için seçilebilir olmaz, dolayısıyla madde ayırt ediciliği düşer. Bu nedenle çeldiricilerin nasıl çalıştığını incelemek önemlidir. Özellikle testi alan bireylerin çok düşük bir kısmı tarafından seçilen çeldiriciler kontrol edilmeli, gerekli görülürse değiştirilmelidir. Ayrıca, özellikle başarısız gruptaki öğrencilerin doğru yanıtlayıp başarılı gruptaki öğrencilerin yanlış yanıtladığı maddelerde çeldiriciler analiz edilip olası sorunlar giderilmelidir.

İyi bir çeldirici başarısız öğrenciler tarafından, başarılı öğrencilere göre daha çok tercih edilmelidir. Bu bakımdan madde ayırt ediciliğinde kullanılan teknikler çeldirici analizi için de kullanılabilir. Şöyle ki alt – üst %27'lik gruplar yönteminde en başarılı %27'lik grubun doğru yanıtlama yüzdesinden, en başarısız %27'lik grubun doğru yanıtlama yüzdesi çıkarılıyordu. Aynı işlem, doğru yanıtlayan yüzdesi yerine bir çeldiriciyi seçme yüzdesi için yapıldığında beklenen durum üst grubun çeldiriciyi seçme yüzdesinin, alt grubunkinden küçük olmasıdır. Bir başka deyişle ideal olan alt gruptaki kişilerin daha yüksek bir oranda çeldiriciye gitmiş olmasıdır. Tahmin edileceği üzere bu durumda çeldiricilere ait istatistiklerin negatif değer alması beklenir (Gronlund, 1977). Eğer pozitif bir değer ortaya çıkıyorsa, yani bir çeldiriciye üst gruptaki öğrenciler, alt gruptaki öğrencilere göre daha fazla yöneliyorsa bu çeldirici ters çalışmış anlamına gelir. Böyle bir sonuçla karşılaşıldığında önce maddede bir hata olup olmadığı kontrol edilmeli, daha sonra çeldirici gözden geçirilmelidir.

Bazen bir çeldirici kimse tarafından seçilmeyebilir. Bu durumda çeldirici çalışmamış demektir. Sınıf içi öğrenci başarısının belirlenmesi amacıyla yapılan bir testte bu tür bir sonuçla karşılaşılması, öğrencilerin maddenin ölçmeye çalıştığı kazanıma ulaştığı şeklinde yorumlanabilir. Ancak çok sayıda öğrencinin / adayın aldığı standart testlerde, her bir çeldiricinin seçilmiş olması istenir. Haladyna ve Downing (1993), testi alan grubun %5'i tarafından seçilmemiş olan bir çeldiricinin, kimse tarafından seçilebilir olmadığını, dolayısıyla bu çeldiricinin çalışmadığını ifade etmiştir. Aynı zamanda bu bölümün Uygulama kısmında gösterilen Şekil 17 gibi bir madde için seçenek yanıt eğrisi elde edildiğinde, beklenen durum başarı arttıkça çeldiricinin seçilme oranının düşmesidir.

3.4. Madde Güvenirlik İndeksi

Madde güvenirlik indeksi, madde standart sapmasının ve madde – toplam korelasyonun çarpımı ile elde edilen bir katsayıdır. Hatırlanacak olursa 1 – 0 puanlanan maddelerde, madde standart sapması $\sqrt{p_j q_j}$ formülüyle hesaplanmaktadır. Bu formülden anlaşılacağı gibi aynı ayırt ediciliğe sahip iki maddeden varyansı (dolayısıyla standart sapması) daha büyük olan, testten alınan puanların güvenirliğine daha çok katkı sağlar. Öte yandan testteki maddeler orta güçlükte – ya da birbirine yakın güçlükte – ise maddelerin standart sapmaları da birbirine yakın olur. Böyle bir durumda test oluşturulurken sadece madde ayırt ediciliğini dikkate alma ile madde güvenirliğini dikkate alma arasında büyük bir fark oluşmaz (Crocker & Algina, 2008).

4. UYGULAMA

Bu bölümde hazır bir veri seti kullanarak test ve madde istatistikleri elde edilecektir. Bunun için başvurulabilecek çeşitli yazılımlar mevcuttur. Bununla beraber açık kaynaklı olması ve kullanım kolaylığı sunması bakımından burada Shiny Item Analysis (Martinkova & Hladka, 2023; Martinkova & Drabinova, 2018) yazılımı tercih edilmiştir. Shiny Item Analysis (SIA), KTK ve MTK temelli analizler, madde kalibrasyonları, değişen madde fonksiyonu, faktör analizi gibi birçok analizi yapabilen bir R paketidir. Bunun yanında Shiny arayüzü sayesinde herhangi bir kodlama bilgisi gerektirmeden kolayca kullanılabilmektedir. Dileyen okuyucular SIA'yı *shinyitemanalysis.org* üzerinden Quick Start (Hızlı Başlat) bağlantısına tıklayarak çalıştırabilir; dileyenler ise R'da:

```
install.packages("ShinyItemAnalysis")
library(ShinyItemAnalysis)
startShinyItemAnalysis()
```

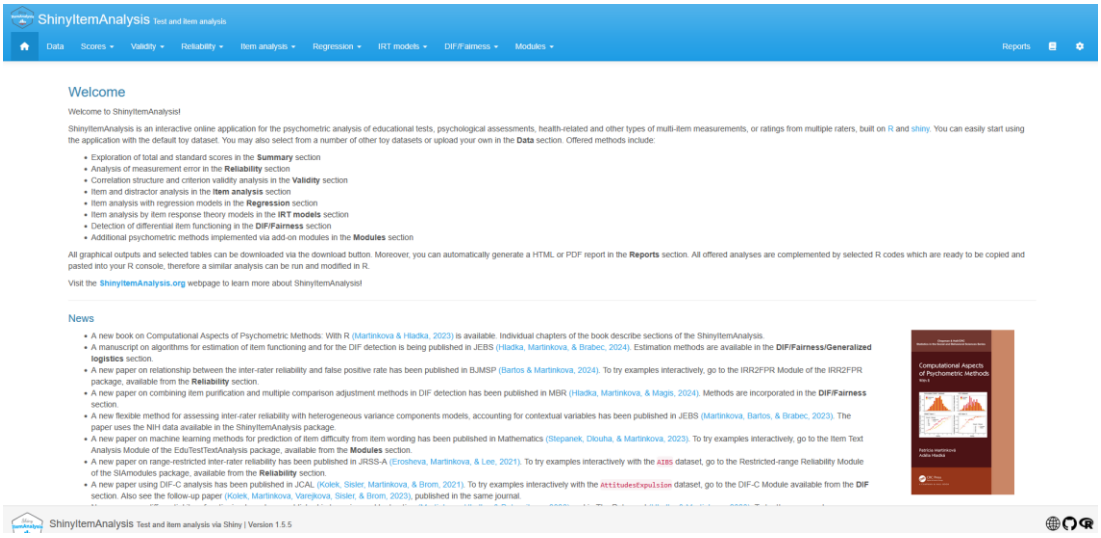
komutlarını kullanarak SIA'yı yükleyebilir ve ardından SIA'yı çağırıp uygulamayı başlatabilir.

4.1. Giriş Ekranı

SIA başlatıldığında Şekil 5'teki gibi bir ekranla karşılaşılır.

Şekil 5

SIA Karşılama Ekranı



Bu ekran bir karşılama ekranı olup SIA'nın kısa bir tanıtımını ve yazılımla ilgili güncel gelişmeleri içerir. Veri yükleme ve analiz işlemlerinin yapılacağı sayfaya ise Şekil 6'da yer alan üst menüden ulaşılabilir.

Şekil 6

SIA Üst Menüsü



Bu menüdeki ögeler işlevleri ile birlikte aşağıda listelenmiştir:

- *Data:* Veri yükleme
- *Scores:* Test istatistikleri (Merkezî eğilim ve dağılım ölçüleri, standart puanlar)
- *Validity:* Ölçüte ve içyapıya dayalı geçerlik kanıtları (korelasyon matrisi, faktör analizi)
- *Reliability:* İç tutarlılık anlamında güvenirlik katsayıları
- *Item analysis:* Madde analizleri (Madde güçlüğü, ayırt ediciliği, madde - yanıt grafikleri)
- *Regression:* Lojistik regresyon (özellikle MTK ile ilişkisini incelemek amacıyla)
- *DIF/Fairness:* Değişen Madde Fonksiyonu, Yanlılık
- *Modules:* Kullanıcılar tarafından eklenen modüller

4.2. Data (Veri) Ekranı

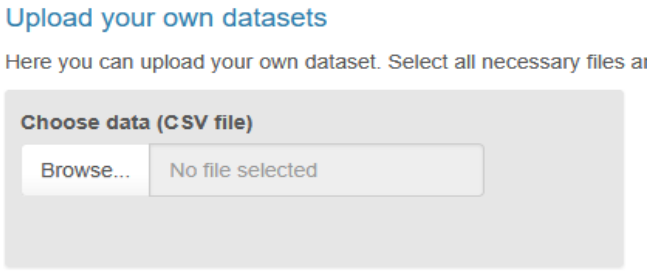
Data sekmesinde; Data (Veri), Basic Summary (Temel özet) ve Data Exploration (Veriyi keşfet) şeklinde üç alt sekme bulunur. Data sekmesinden kullanılacak veriler seçilir/yüklenir. Basic summary sekmesinde verilere ait temel özet bilgiler (en küçük ve en büyük değerler, kişi sayısı gibi) sunulur. Data exploration sekmesinde ise yüklenen veriler incelenebilir. Böylece doğru veri ile çalışılıp çalışılmadığı kontrol edilir.

Data alt sekmesinde kullanıcılar kendi verilerini yükleyebilecekleri gibi yazılım içerisindeki hazır veri setlerinden de yararlanabilirler. Tüm okuyucuların aynı okuma deneyimini elde edebilmesi adına bu bölümde yazılım içerisinde yer alan hazır verilerden GMAT verisi ile çalışılmıştır. Ancak aşağıda kendi verisini yüklemek isteyen okurların bunu nasıl yapabileceği de açıklanmıştır.

Çalışılacak veriyi yüklemek için "Upload your own datasets" (kendi veri setinizi yükleyin) başlığı altındaki "Choose data" (Veri seç) bölümü kullanılır (Şekil 7). SIA, yalnızca CSV formatındaki verilerle çalışmaya izin verdiğinden verinin öncelikle bu formatta kaydedilmesi gerekir. Microsoft Excel gibi birçok yazılım verinin CSV formatında kaydedilmesine izin vermektedir.

Şekil 7

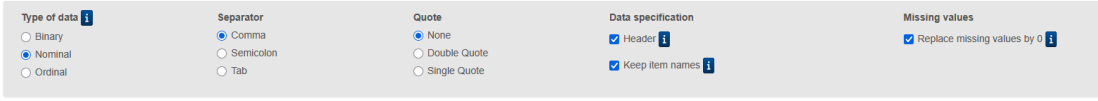
Veri Yükleme Alanı



Çalışılacak veri seçildikten sonra veri yapısına ilişkin bazı bilgilerin tanımlanması gerekir. Bunun için Şekil 8'deki bölümden tercihler yapılmalıdır.

Şekil 8

SIA Veri Dosyası Özelliklerinin Tanımlanması



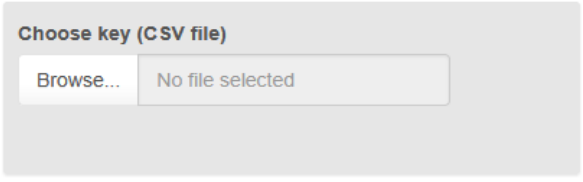
Bu bölümde kullanıcının tercih edebileceği seçenekler ve anlamları şöyledir:

- *Type of data (veri tipi):* Nasıl bir veri yüklendiği seçilir. 1 - 0 (binary), kategorik yani A-B-C-D gibi (nominal) ya da 1-2-3-4-5 gibi sıralama (ordinal) seçeneklerinden biri işaretlenmelidir.
- *Seperator (ayraç):* CSV dosyasında sütunların nasıl ayrıldığının seçilmesi gerekir. Virgül (comma), noktalı virgül (semicolon) ya da sekme (tab) seçeneklerinden biri işaretlenmelidir. CSV dosyasının nasıl ayrıldığından emin olunamaması halinde, CSV dosyası not defteri ile açılarak kontrol sağlanabilir. Microsoft Excel varsayılan olarak noktalı virgül ayracı ile CSV dosyası oluşturmaktadır.
    - *Quote (tırnak işareti):* Veride metin varsa bu metinlerin veride nasıl belirtildiğinin girildiği bölümdür. Verideki metin "Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı" şeklindeyse Single Quote yani tek tırnak; ""Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı"" şeklindeyse Double Quote yani çift tırnak seçilmelidir. Metinlerin veride nasıl yer aldığı bilinmiyorsa CSV dosyası not defteri benzeri bir uygulama ile açılarak kontrol edilebilir.
- *Data specification (veri özellikleri):* Verinin ilk satırı değişken isimlerinden oluşuyorsa "Header" seçili olmalıdır. Değişken isimlerinin aynen korunması isteniyorsa "Keep item names" seçeneği seçili kalmalıdır, aksi halde SIA tarafından otomatik olarak değişken ismi atanacaktır.
- *Missing values (kayıp değerler):* Verideki kayıp değerlerin sıfır olarak puanlanması / kodlanması isteniyorsa "Replace missing values by 0" seçili olmalıdır.

Bir sonraki girdi seçeneği Şekil 9'da gösterilen ve cevap anahtarının yüklenebileceği bölümdür.

Şekil 9

Cevap Anahtarı Yükleme Alanı

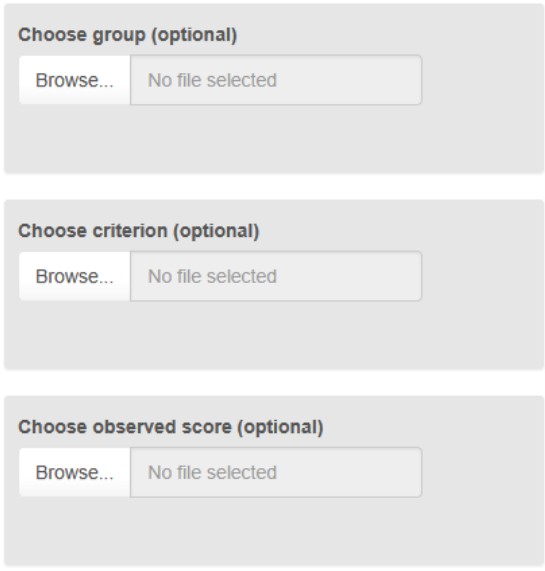


Eğer yüklenen veri 1 – 0 olarak puanlanmış veri değil de doğrudan A - B - C - D gibi yanıtlardan oluşan bir veriyse bu bölüme cevap anahtarının yüklenmesi gerekir. SIA cevap anahtarına göre otomatik olarak puanlamayı yapar. Ancak A ve B kitapçığı gibi bir desteği bulunmamaktadır. Bu nedenle yüklenecek verilerin tamamının aynı kitapçığa göre düzenlenmiş olması şarttır. Ayrıca A - B - C - D gibi kategorik bir veri ve cevap anahtarı yüklenecekse bir üstteki Type of data kısmından Nominal seçili olmalıdır.

Daha sonra Değişen Madde Fonksiyonu / Yanlılık gibi çalışmalarda kullanmak üzere grup verisinin; ölçüte dayalı geçerlik çalışmasında kullanmak üzere ölçüt puanların ve gözlenen puanların yüklenebileceği alan gelmektedir (bkz. Şekil 10).

Şekil 10

Grup, Ölçüt ve Gözlenen Puan Yükleme Alanı



Bu verilerin yüklenmesi zorunlu olmayıp ihtiyaç duyulması halinde yüklenir. Her bir verinin CSV türünde olması gerektiği unutulmamalıdır. Yüklenmek istenen veriler seçildikten sonra sağ alttaki Upload data (Veriyi yükle) düğmesine tıklanarak veriler yüklenebilir.

Bu bölümde örnek olarak GMAT verisi ile çalışılacağı belirtilmişti. Dolayısıyla "Training datasets" altından GMAT verisi seçilerek uygulamanın sonraki aşamalarına devam edilmiştir.

GMAT verisinin nasıl bir yapıda olduğunu görmek için Data sekmesi altından sırasıyla Basic summary ve Data exploration alt sekmeleri incelenir. Basic summary alt sekmesinde; sırasıyla ana veri seti, puanlanmış veri seti, grup, ölçüt değişken ve gözlenen puanlara ait bir özet sunulur. Data exploration alt sekmesinde ise doğrudan verilerin ham hali incelenebilir. Böylece özellikle kendi verisini yükleyen kullanıcılar verinin doğru bir şekilde yüklenip yüklenmediğini görebilirler.

4.3. Test İstatistikleri

Test puanlarına ait aritmetik ortalama, ortanca, standart sapma gibi değerlerin yanında standartlaştırılmış z ve T-puanlarına ait bilgileri elde etmek için Scores (puanlar) sekmesinden yararlanılır. Scores altında Total scores (toplam puanlar) ve Standard scores (standart puanlar) olmak üzere iki seçenek yer alır. Total scores seçildiğinde açılan sayfada Şekil 11'de görülen temel test istatistikler elde edilebilir.

Şekil 11

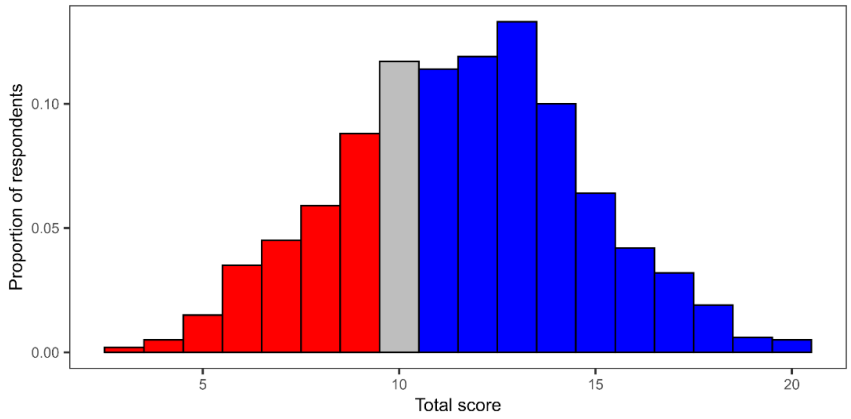
Temel Test İstatistikleri

n	n_c	Min	Max	Mean	Median	SD	Skewness	Kurtosis
2000	2000	3.00	20.00	11.60	12.00	3.11	0.00	2.75

Burada öncelikle çalışılan verinin kaç kişiden oluştuğu (n), kayıp değerler çıkarıldıktan sonra kaç kişinin verisinin yer aldığı (n_c), en küçük değer (min), en büyük değer (max), aritmetik ortalama (mean), ortanca (median), standart sapma (SD), çarpıklık (skewness) ve basıklık (kurtosis) değerleri yer almaktadır. Örnekte 2000 kişilik bir veriyle çalışıldığı, en küçük puanın 3 ve en büyük puanın 20 olduğu görülmektedir. Ortalama 11,6 ve ortanca 12 olarak elde edilmiştir. Standart sapma ise 3,11'dir. Ortalama ve ortancanın birbirine yakınlığı, çarpıklığın sıfıra çok yakın olması; basıklığın ise 3 civarında olması nedeniyle puanların normale yakın olduğu söylenebilir. Bu durum tablonun hemen altında yer alan histogramla da desteklenmektedir (bkz. Şekil 12).

Şekil 12

Test Puanlarına İlişkin Histogram



Histogramdaki gri alan kesme noktasını, kırmızı alan kesme noktasının altını, mavi alansa kesme noktasının üstünde puan alanları temsil etmektedir. Kesme noktası kullanıcı tarafından histogramın hemen üzerinde yer alan kaydırma çubuğundan ayarlanabilmektedir.

Standard scores menüsüne gelindiğinde ise her bir ham puanın (total score) sırasıyla yüzdelik dilimi (percentile), yüzlük puanı (success rate), z ve T-puanları görülebilmektedir (bkz. Şekil 13). Şekil 13'te görüldüğü gibi SIA standart puanları her bir öğrenci için ayrı ayrı vermek yerine, her bir ham puana karşılık gelen standart puanları raporlamaktadır.

Şekil 13

Test Puanlarına İlişkin Standart Puanlar

Table by score

Total score	Percentile	Success rate	Z-score	T-score
3.00	0.00	15.00	-2.77	22.30
4.00	0.01	20.00	-2.45	25.52
5.00	0.02	25.00	-2.13	28.74
6.00	0.06	30.00	-1.80	31.96
7.00	0.10	35.00	-1.48	35.18
8.00	0.16	40.00	-1.16	38.40
9.00	0.25	45.00	-0.84	41.62
10.00	0.37	50.00	-0.52	44.84
11.00	0.48	55.00	-0.19	48.06
12.00	0.60	60.00	0.13	51.28
13.00	0.73	65.00	0.45	54.50
14.00	0.83	70.00	0.77	57.72
15.00	0.90	75.00	1.09	60.94
16.00	0.94	80.00	1.42	64.16
17.00	0.97	85.00	1.74	67.38
18.00	0.99	90.00	2.06	70.60
19.00	1.00	95.00	2.38	73.82
20.00	1.00	100.00	2.70	77.04

Validity (geçerlik) menüsü üzerinden ölçüt geçerliği ve korelasyon matrisi ile içyapıya dayalı geçerlik kanıtına (faktör analizi aracılığıyla) ulaşılabilmektedir. Ölçüte dayalı geçerlik kanıtı elde edebilmek için Data sekmesinde "Choose criterion" kısmına veri yüklenmiş olmalıdır. Örnekte her bir öğrenci için ölçüt puanlar da girilmiş olduğundan *Criterion Validity* menüsüne girildiğinde bir saçılım grafiği ve ardından ölçüt ile test puanları arasındaki korelasyon katsayısı raporlanmaktadır. Correlation structure kısmında ise maddeler arası korelasyon katsayıları raporlanmakta, aynı zamanda kolay yorumlama için sıcaklık grafiği (heatmap) sunulmaktadır. Korelasyon katsayısı olarak Pearson, Spearman ya da Polychroic alternatiflerinden biri seçilebilmektedir. Maddeler 1-0 şeklinde puanlanıyorsa tetrakorik korelasyon katsayı kullanılmalı, bunun için Polychroic seçeneği işaretlenmelidir. *Factor analysis* menüsünde ise testin nasıl bir faktör yapısını ölçtüğüne dair inceleme yapılabilmektedir. Bu menüde temel bileşen analizi (PCA) veya faktör analizi (FA); Pearson ya da polikorik korelasyon matrisine dayalı olarak yapılabilmektedir. Faktör analizi ve korelasyon katsayılarının detaylı sunumu bu bölümün kapsamına girmediğinden ilgili kısmın anlatımı sınırlandırılmıştır.

Reliability (güvenirlik) menüsünde ise iç tutarlılık anlamındaki güvenirlik katsayıları elde edilebilmektedir. Bu amaçla Spearman-Brown tahmini, iki yarıya bölme yöntemi ve Cronbach alfa katsayıları hesaplanmaktadır. Spearman-Brown tahmininde "Test X sayıda benzer maddeden oluşsaydı güvenirlik katsayısı ne olurdu?" ya da "Güvenirlik katsayısının Y olması isteniyorsa test buna benzer kaç maddeden oluşmalıydı?"

sorularına yanıt aranmaktadır. Örnek veride 20 madde bulunmakta ve güvenirlik katsayısı 0,55 olarak hesaplanmaktadır. *"Number of items in new data"* (Yeni verideki madde sayısı) kısmına 30 yazılırsa hemen altında 0,647 değeri görülür. Bunun anlamı, teste benzer nitelikte 10 madde daha eklenirse, yani testin toplamda 30 maddeden oluşursa güvenirlik katsayısının 0,647 olması beklenmektedir. Hemen alttaki "Reliability of new data" (Yeni verinin güvenirliği) kısmına ise 0,80 yazılırsa, hemen altında 66 madde olarak hesaplama yapılır. Yani 0,80 güvenirlik katsayısına ulaşılması bekleniyorsa, testtekilere benzer 46 madde daha teste eklenmeli ve toplam 66 maddelik bir test oluşturulmalıdır. *Split-half method* (iki yarıya bölme yöntemi) menüsünde ise iki yarıya bölme yöntemine göre güvenirlik katsayısı hesaplanmaktadır. Bu menüde ilk yarı – son yarı, tek – çift, rastgele, ortalama ve en kötü seçenekleri bulunmaktadır. Ayrıca 10000 farklı şekilde test iki yarıya bölünerek her birinin sonucundan bir histogram oluşturularak raporlanmaktadır. *Cronbach alfa* menüsünde ise teste ait Cronbach alfa katsayısı hesaplanmaktadır. 1 – 0 puanlanan maddelerde Cronbach alfa, aynı veri seti için hesaplanacak KR-20 katsayısına eşit olacağından sayfada KR-20 için ayrı bir menüye yer verilmemiştir. Güvenirlik ve geçerlik kanıtları, faktör analizinin yapılması gibi konular bu bölümün kapsamı dışında kaldığından ilgili menüler ile ilgili açıklamalar bu kadarla sınırlandırılmıştır.

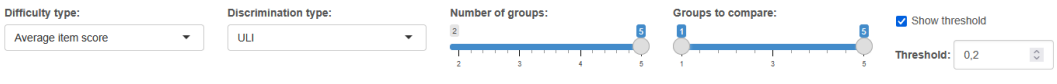
4.4. Madde Analizi

SIA ile madde analizleri temelde *Item Analysis* (madde analizi) menüsünden yapılmaktadır. Bunun yanında *DIF/Fairness* (değişen madde fonksiyonu / yanlılık) menüsünden maddelerin farklı gruplarda benzer çalışıp çalışmadığı incelenebilmektedir.

Item Analysis menüsü altındaki ilk seçenek *Traditional Item Analysis'dir* (geleneksel madde analizi). Bu menü altında KTK temelli madde güçlüğü ve ayırt ediciliğine ilişkin istatistikler elde edilebilmektedir. Şekil 14'te yer alan seçenekler aracılığıyla Şekil 15'te verilen grafiğe ilişkin ayarlamalar yapılabilmektedir.

Şekil 14

Madde Analizi Seçenek Alanı



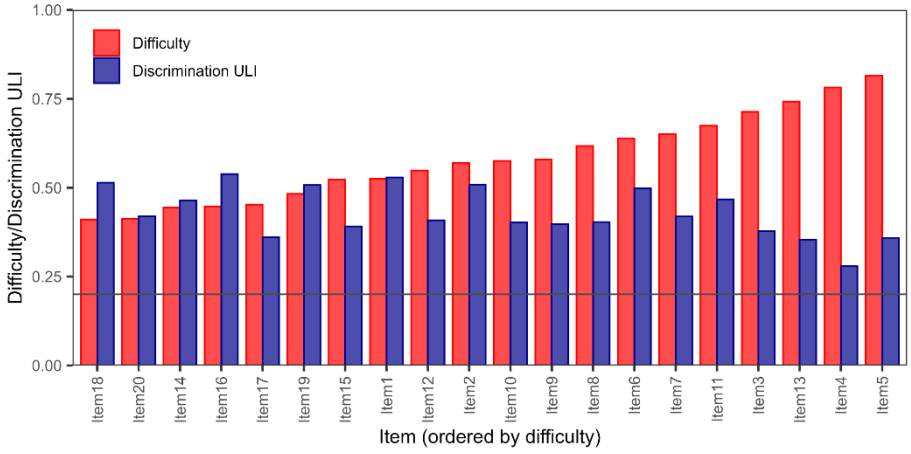
Difficulty type (güçlük tipi) kısmında iki seçenek bulunmaktadır: *Average scaled score*, madde puan ortalamasının madde ranjına bölünmesiyle elde edilen ölçeklenmiş puanı verirken *Average item score* maddenin ortalama puanını vermektedir. Çok kategorili puanlanan maddeler (örneğin, açık uçlu maddeler) için average scaled score 0 - 1 aralığında madde güçlüğünü sunmaktadır. İki kategorili puanlanan maddeler içinse her iki seçenek aynı sonucu vermektedir.

Discrimination type (ayırt edicilik tipi) kısmında ise dört seçenek yer almaktadır. *ULI: Upper-Lower Index*, alt-üst gruplar yöntemi ile elde edilen indekstir. *RIT*, madde – toplam korelasyonunu; *RIR* ise madde – kalan korelasyonunu verir. *None* seçeneği seçildiğinde ise madde ayırt ediciliği raporlanmaz. *Number of groups* ve *groups to compare* seçenekleri alt – üst gruplar yöntemi seçildiğinde aktif hale gelir. *Number of groups*, verinin kaç gruba bölüneceğini; *groups to compare* ise ayırt edicilik indeksinin hangi iki grubun karşılaştırılmasına göre hesaplanacağını belirler. Dolayısıyla her ne kadar bu indeks alt – üst gruplar yöntemi olarak anılsa da örneklem beş gruba bölünüp maddenin dördüncü ve beşinci grubu birbirinden ayırıp ayırmadığı incelenebilmektedir.

Son olarak Şekil 15'teki grafiğe yatayda bir kesme çizgisi eklemek için *Show threshold* seçeneği işaretlenerek alttaki kutucuktan kesme çizgisinin hangi değerde olması gerektiği belirtilir. Tüm bu ayarlamalardan sonra Şekil 15'teki gibi bir grafik elde edilebilir.

Şekil 15

Madde Analizi Sonuçları – I



Şekildeki grafik maddeleri güçlüklerine göre (kırmızı renkli çubuklar) sıralamakta, aynı zamanda ayırt edicilik indeksini de (mor çubuklar) göstermektedir. Bu grafiğin altında ise Şekil 16'daki gibi bir tablo yer almaktadır.

Şekil 16

Madde Analizi Sonuçları – II

	Diff.	Avg. score	SD	Min.	Obs. min.	Max.	Obs. max.	Prop. max.	RIT	RIR	I-C cor.	ULI	Rel.	Val.	α-drop	Missed [%]	Not-reached [%]
Item1	0.52	0.52	0.50	0	0	1	1	0.52	0.38	0.23	0.09	0.43	0.19	0.05	0.53	0.00	0.00
Item2	0.57	0.57	0.50	0	0	1	1	0.57	0.37	0.22	0.09	0.42	0.18	0.04	0.53	0.00	0.00
Item3	0.71	0.71	0.45	0	0	1	1	0.71	0.27	0.13	0.09	0.29	0.12	0.04	0.55	0.00	0.00
Item4	0.78	0.78	0.41	0	0	1	1	0.78	0.25	0.12	0.12	0.25	0.10	0.05	0.55	0.00	0.00
Item5	0.81	0.81	0.39	0	0	1	1	0.81	0.33	0.21	0.10	0.29	0.13	0.04	0.54	0.00	0.00

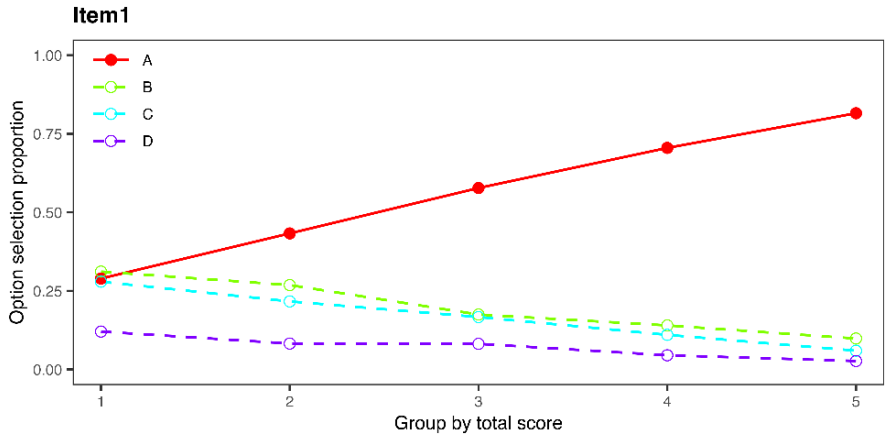
Bu tablodaki *Diff.* madde güçlüğünü; *Avg. score* ortalama puanı temsil etmektedir. 1-0 puanlanan maddeler için *Diff.* ve *Avg. score* sütunları aynı değere sahip olur. *SD*, maddenin standart sapmasını; *min.* o maddeden alınabilecek en düşük puanı; *Obs. min*, o maddeden alınan en düşük puanı; *max.* alınabilecek en yüksek puanı; *Obs. max* alınan en yüksek puanı göstermektedir. *Prop. max.* ise maddeden en yüksek puanı alan kişilerin oranını göstermektedir. 1-0 puanlanan maddeler için *Prop. max.* madde güçlüğünü bildirmektedir.

RIT, madde - toplam korelasyonunu; *RIR* ise madde – kalan korelasyonunu göstermektedir. *I-C cor.* ise madde ile ölçüt puan arasındaki korelasyonu, *ULI*: Alt - üst gruplar yöntemine göre ayırt edicilik indeksini göstermektedir. *Rel.* madde güvenirliğini; *Val.* ise madde geçerlik indeksini vermektedir. *α-drop*, madde çıkarılırsa Cronbach alfa katsayısının nasıl değişeceğini, *Missed [%]* kayıp değer yüzdesini ve son olarak *Non-reached [%]* ise bu maddeye ulaşamayan öğrencilerin yüzdesini temsil etmektedir.

Yine *Item Analysis* menüsü altında yer alan *Item Response Curves* (madde tepki eğrileri) yolu izlenirse testi alan bireylerin toplam puanlarına göre üç gruba ayrıldığı ve her seçeneğin bu gruplarda hangi oranda seçildiği bir grafik görüntülenir. Grup sayısı ayarından, grup sayısı üç yerine beş yapıldığında Şekil 17'deki gibi bir grafik elde edilir.

Şekil 17

Seçeneklerin Analizi



Bu grafikte çeldiriciler kesikli çizgi, anahtar yanıt (doğru cevap) ise düz çizgi ile temsil edilmektedir. Grafik incelenerek çeldiricilerin ve doğru yanıtın farklı gruplarda seçilme oranları görülebilmektedir. Buna göre grup daha başarılı hale geldikçe çeldiricilerin seçilme oranı azalmakta ve anahtar yanıtın seçilme oranı artmaktadır. Her seçenek öğrenciler tarafından belli oranlarda işaretlenmiş olup çalışmayan çeldirici yoktur. Yalnızca D çeldiricisi en başarısız grup tarafından bile diğer çeldiricilere göre daha az tercih edilmiştir.

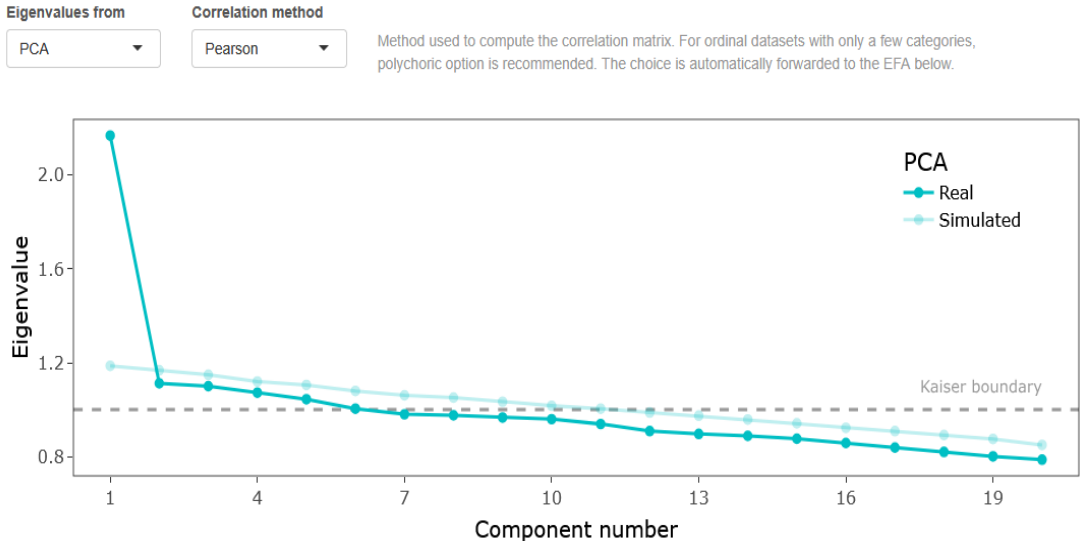
4.5. MTK Analizleri

Varsayımların Kontrolü

Tek boyutlu MTK modelleri ile çalışıldığında ilk varsayım maddelerin tek bir başat faktörü ölçüyor olmasıdır. Bunu kontrol etmek için faktör analizinden yararlanılabilir. Faktör analizi bölümün kapsamı dışında kaldığından burada yalnızca yamaç birikinti grafiği kullanılarak tek boyutluluk incelenecektir. Bunun için *Validity* menüsü altındaki *Factor analysis* yolu kullanılacaktır. Şekil 18'deki gibi açılan sayfada varsayılan olarak *Temel Bileşenler Analizi (Principal Component Analysis-PCA)* ve maddeler arası korelasyon katsayısı için Pearson momentler çarpım korelasyon katsayısı seçilidir.

Şekil 18

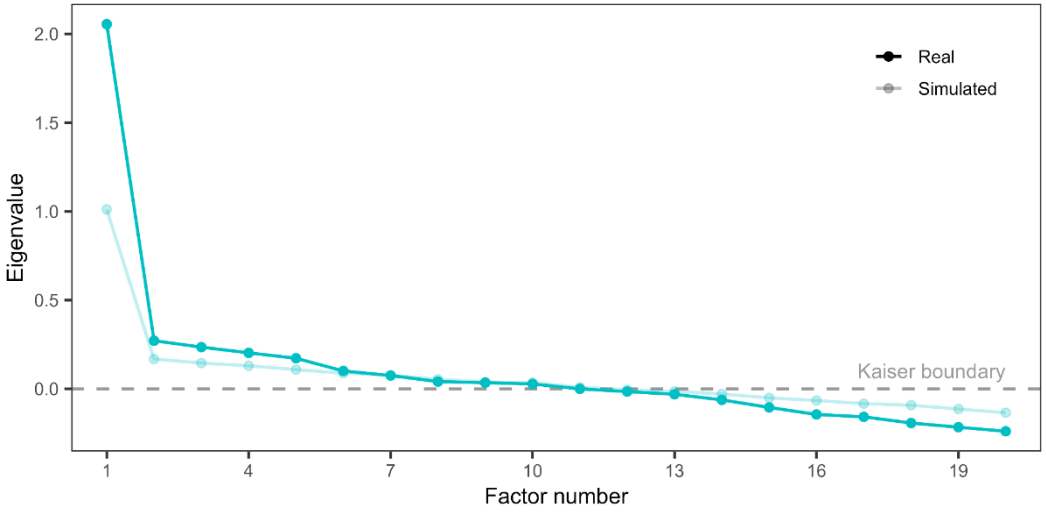
Temel Bileşen Analizine Göre Yamaç-Birikinti Grafiği



Ancak psikolojik ölçmelerde, maddeler ile altta yatan gizil bir yapı ortaya konulmaya çalışıldığından temel bileşen analizi yerine faktör analizinin yapılması tercih edilir. Ayrıca örnekteki veri 1-0 şeklinde puanlandığı için korelasyon yöntemi olarak tetrakorik korelasyon katsayısı kullanılmalıdır. Bu amaçla Eigenvalues from (özdeğerler) ve Correlation method (korelasyon yöntemi) kutucuklarında sırasıyla FA ve Polychoric ifadeleri seçilmelidir. Bu seçimlerin ardından yamaç birikinti grafiği otomatik olarak Şekil 19'daki gibi güncellenir.

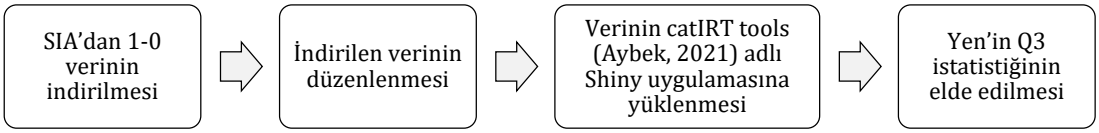
Şekil 19

Faktör Analizine Göre Yamaç-Birikinti Grafiği



Bu yamaç birikinti grafiğinde Real (gerçek) ile gösterilen, gerçek veriden elde edilen öz değerler iken Simulated (simülatif) ile gösterilen paralel analiz sonucudur. Her ikisine göre de yirmi maddenin tek bir başat faktör altında toplanabildiği anlaşıldığından tek boyutluluk varsayımının sağlandığı söylenebilir. İlgisi olan okurlar bu grafiğin altındaki tabloları inceleyerek bütün faktör analizini tamamlayabilirler.

Her ne kadar tek boyutluluk varsayımı karşılandığında yerel bağımsızlığın sağlandığı yönünde karar verilebileceğini belirten kaynaklar olsa da yerel bağımsızlığın testinde Yen'in Q3 istatistiği sıklıkla kullanılmaktadır. Ne yazık ki SIA, henüz Yen'in Q3 istatistiğini vermemektedir. Bu nedenle Yen'in Q3 istatistiği farklı bir Shiny uygulaması ile incelenecektir. Yen'in Q3 istatistiğini elde etmek için sırasıyla aşağıdaki adımlar izlenecektir:



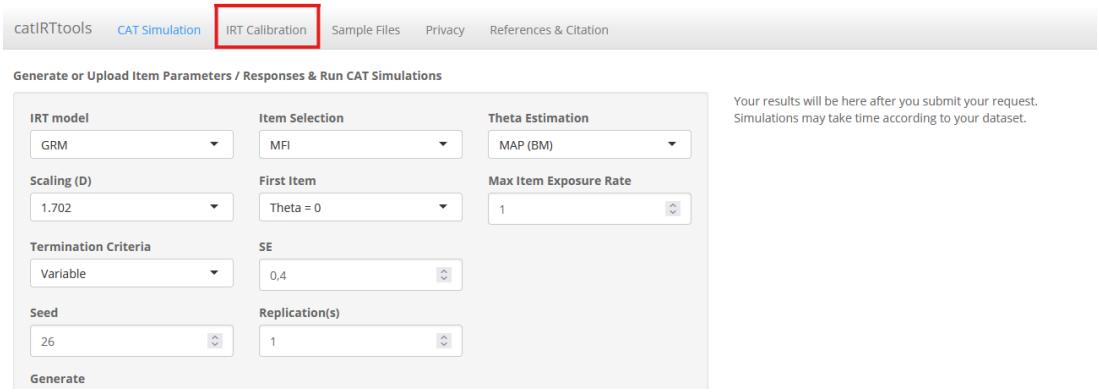
Yen'in Q3 istatistiği için başvurulacak olan catIRT tools uygulamasına https://eptlab.com/catirt adresinden çevrimiçi erişim sağlanabilir ya da https://github.com/ecaybek/catIRTtools adresinde yer alan kaynak kodlar indirilerek uygulamanın kullanıcı bilgisayarında da çalıştırılması mümkündür.

Öncelikle SIA'da Data (veri) sekmesinden Data exploration (veriyi keşfet) alt sekmesine tıklanmalıdır. Ardından Scored / binarized data (puanlanmış veri) başlığı altındaki Download scored dataset (puanlanmış veri setini indir) butonuna tıklanır. Bu butona tıklandığında 1-0'a dönüştürülmüş veri .csv formatında bilgisayara indirilir.

catIRT tools uygulaması açıldığında Şekil 20'deki gibi bir ekran görüntüsü ile karşılaşılacaktır.

Şekil 20

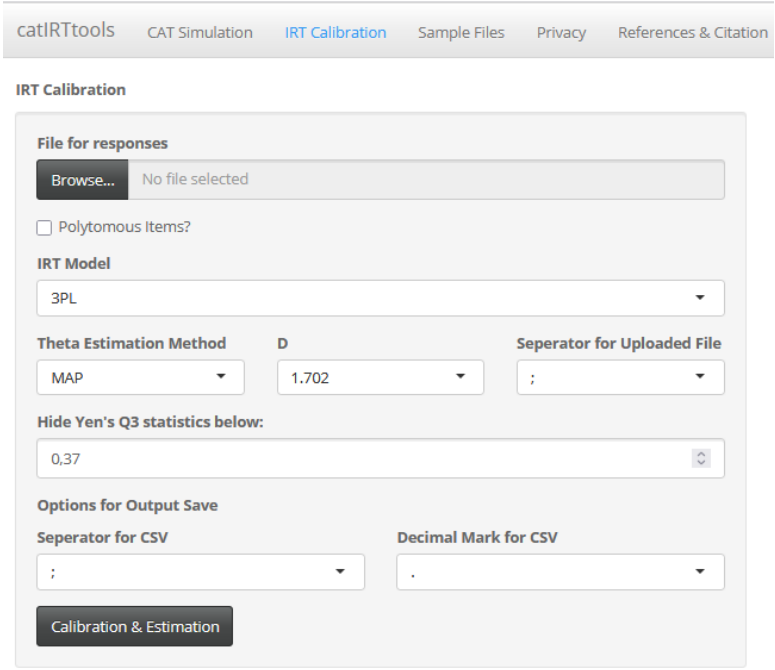
catIRT Tools Ana Ekranı



Uygulamanın ana sayfası, kullanıcıların Bilgisayar Ortamında Bireye Uyarlanmış Test simülasyonu yapmalarına olanak sağlamaktadır. Yen'in Q3 istatistiği için Şekil 20'de kırmızı ile gösterilen IRT Calibration (MTK Kalibrasyonu) sekmesine gidilmelidir. Bu sekmeye ait ekran görüntüsü Şekil 21'de verilmiştir.

Şekil 21

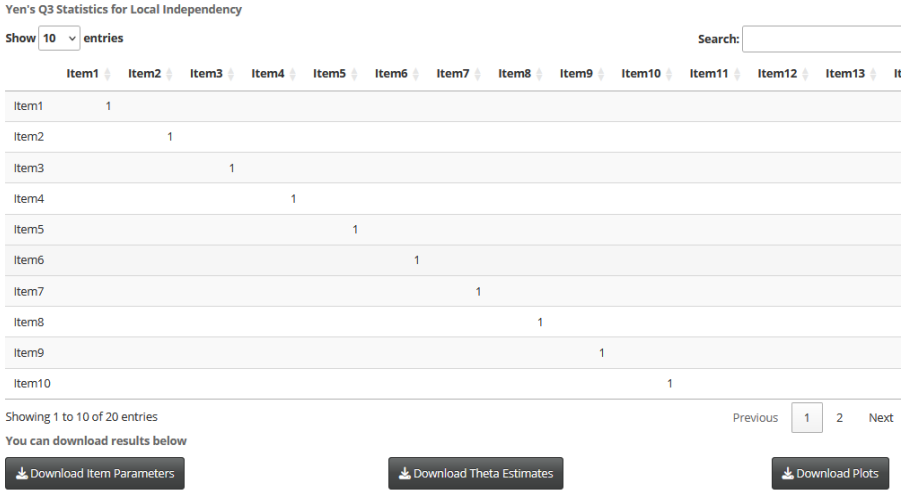
MTK Kalibrasyon Sekmesi



SIA'dan indirilen csv dosyası File for responses (Yanıtlar için dosya) alanına atanmalıdır. Örnekteki dosya 1-0 puanlanan bir veri olduğu için Polytomous items (çok kategorili puanlanan madde) kısmına herhangi bir tik atılmamalıdır. IRT Model kısmından kullanılması planlanan MTK modeline ilişkin seçim yapılmalıdır. Theta estimation method (Teta kestirim yöntemi) kısmından kullanılması planlanan yetenek kestirim yöntemi; D kısmından ölçekleme sabiti (1 ya da 1,702) ve yüklenen csv dosyasında sütunların hangi işaretle ayrıldığı (Seperator for Uploaded File) seçilmelidir. SIA'dan indirilen csv dosyası, sütun ayracı olarak virgül (,) kullandığı için bu bölümde virgül seçilmelidir. Son olarak Hide Yen's Q3 statistics below (Aşağıdakinden küçük Q3 istatistiklerini gizle) kısmına Yen'in Q3 istatistiği için arzu edilen kesme değeri girilmelidir. Bu değer girilmezse tüm madde çiftleri için Yen'in Q3 istatistiği raporlanacağından tabloyu yorumlamak güçleşir. Ölçüt olarak genellikle .20 kullanıldığından bu kısma 0,20 değeri girilebilir. Daha sonra Calibration & Estimation (Kalibre et & kestirim yap) butonuna tıklanır. Sonuçların elde edilmesi verinin büyüklüğüne göre zaman alıcı olabilir. Sonuçlar elde edildiğinde Şekil 22'deki gibi bir tablo ve butonlar görünür.

Şekil 22

Yen'in Q3 İstatistikleri



İlk gösterilen tabloda madde çiftleri arasındaki Yen'in Q3 istatistikleri raporlanmaktadır. Analiz öncesinde 0,20 değerinin altındaki Q3 istatistiklerinin gösterilmemesi istendiği için madde çiftleri arasında herhangi bir değer görülmemektedir. Her maddenin kendisiyle olan Q3 istatistiği ise 1 olarak raporlanmaktadır, ancak bu dikkate alınmamalıdır. Ayrıca en alttaki butonlardan sırasıyla madde parametreleri, bireylere ait yetenek kestirimleri ve madde karakteristik eğrileri indirilebilir. Ancak SIA, MTK bağlamında daha çok özellik sunduğu için bu uygulamadan yalnızca yerel bağımsızlık kontrolü için yararlanılmıştır. Sonuç olarak hiçbir madde çiftinin yerel bağımsızlığı bozmadığına karar verilmiş ve MTK analizlerine SIA üzerinden devam edilmiştir.

Model Karşılaştırma

Verinin hangi MTK modeli ile en yüksek uyumu gösterdiğini belirlemek için SIA'da IRT models (MTK modelleri) sekmesi altındaki Model comparison (model karşılaştırma) yolu izlenir. Açılan sayfada Şekil 23'te yer alan karşılaştırma tablosu incelenir.

Şekil 23

MTK Modeli Karşılaştırma Sonuçları

Table of comparison statistics

Row **BEST** indicates which model has the lowest value of given information criterion.

	AIC	BIC	logLik
1PL	51254.424	51372.043	-25606.212
2PL	51192.115	51416.151	-25556.057
3PL	51200.297	51536.351	-25540.148
4PL	51232.136	51680.209	-25536.068
BEST	**2PL**	**1PL**	

Bu tabloda Akaike bilgi ölçütü (Akaike Information Criterion - AIC), Bayes bilgi ölçütü (Bayesian Information Criterion - BIC) ve logLikelihood değerleri raporlanmaktadır. Ayrıca AIC ve BIC ölçütlerine göre veri ile en yüksek uyumu gösteren model BEST satırında gösterilmektedir. Örnekteki bulgulara göre AIC'e göre 2 PLM; BIC'e göre 1 PLM veri ile en yüksek uyumu göstermektedir. Bu noktada hangi modelle analize devam edileceği test geliştiricisine bağlıdır. Test geliştiricisi aşağıdaki yollardan biri ile ilerlemeyi seçebilir:

1. Madde ayırt edicilik parametrelerini önemli görerek 2PLM ile devam edebilir
2. Daha az sayıda parametreyi değişken tutarak daha genellenebilir bulgular elde etmek için 1PLM ile devam edebilir
3. Her ikisine göre de analizleri yaparak en uygun modele daha sonra karar verebilir.

Uygulamada hem *a* hem de *b* parametrelerinin nasıl değiştiğinin gözlenebilmesi için bu bölümde 2PLM ile devam edilmiştir. Ancak yukarıdaki gerekçelerle 1PLM ile de devam edilmesi yanlış değildir. İki parametreli lojistik modele göre madde kalibrasyonlarının yapılabilmesi için IRT models sekmesi altındaki Dichotomous models (İki kategorili puanlanan modeller) yolu izlenir. Varsayılan olarak Summary (özet) alt sekmesi açılacaktır. Bu sayfada Şekil 24'teki gibi iki adet seçenek sunulmaktadır.

Şekil 24

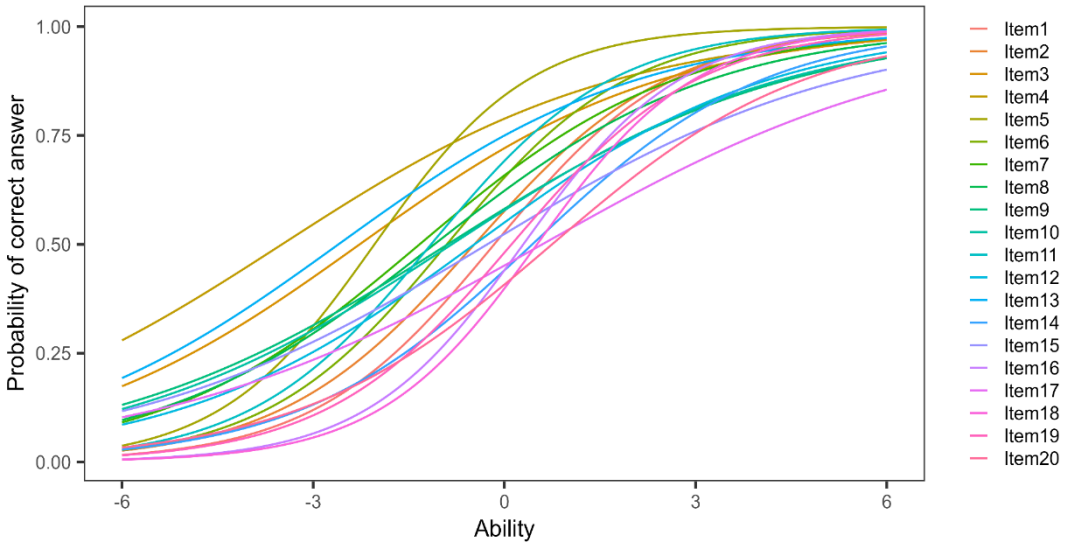
MTK Modeli Seçim Alanı



Model seçeneğinden, maddelerin ölçeklenmesinin istendiği MTK modeli seçilmelidir. Örnek uygulama için bu kısım 2PL olarak seçilmiştir. Parametrization kısmı ise IRT olarak kalmalıdır. Bu sayfa bir özet sayfası olduğu için tüm maddelere ait madde karakteristik eğrileri tek bir grafik üzerinde yer almaktadır. Benzer şekilde tüm maddeler için madde bilgi fonksiyonları da tek bir grafikte sunulmaktadır. Ayrıca test bilgi fonksiyonu ve standart hata değeri de bir grafikte verilmektedir. Tüm bu grafikler sırasıyla Şekil 25 ile Şekil 27 arasında sunulmuştur.

Şekil 25

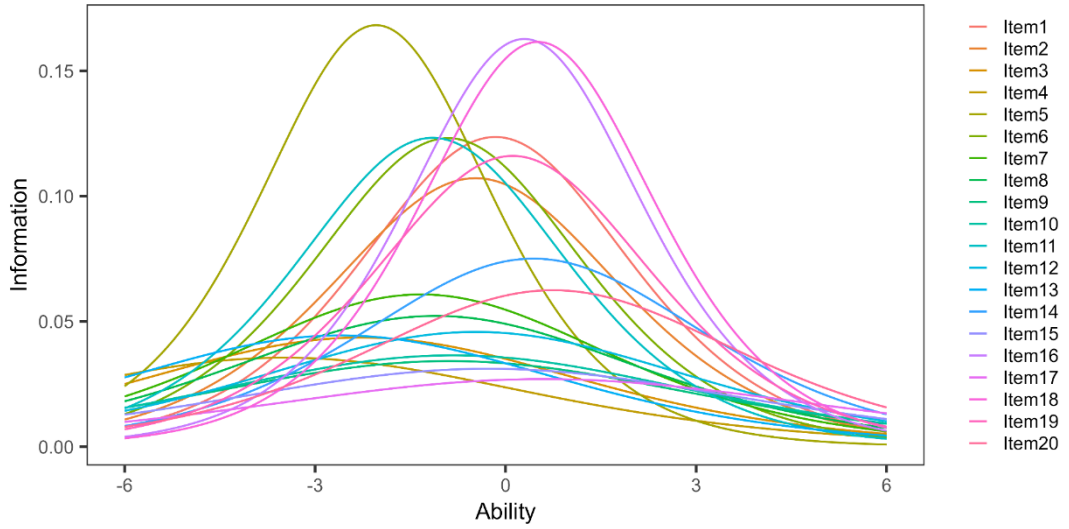
Tüm Maddelere Ait Madde Karakteristik Eğrileri



Bu grafikte her bir maddeye ait madde karakteristik eğrileri yer almaktadır. 2PLM kullanıldığı için tüm maddelerin *a* parametreleri birbirinden farklıdır, bu nedenle her bir eğri farklı eğime sahiptir. Grafiğin x ekseni bireylere ait yetenek düzeyini; y ekseni ise buna karşılık gelen doğru yanıtlama olasılığını temsil etmektedir.

Şekil 26

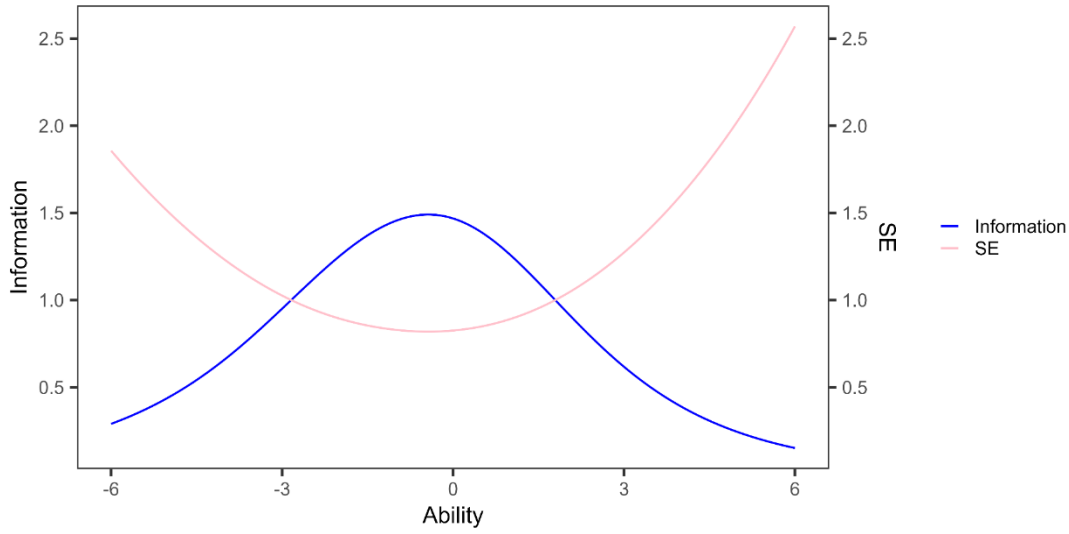
Tüm Maddelere Ait Madde Bilgi Fonksiyonları



Bu grafikte ise maddelere ilişkin bilgi fonksiyonları yer almaktadır. Grafiğin x ekseni yetenek düzeyini gösterirken y ekseni buna karşılık gelen bilgi miktarını temsil etmektedir. Maddenin belli bir yetenek düzeyi için sağladığı bilgi miktarı, maddeye ait *a* parametresinin karesi ile doğrudan ilişkilidir. Haliyle, *a* parametresi küçük olan maddelerin sağladığı bilgi miktarı da düşük olmaktadır. Bu durum Şekil 26'da açık bir şekilde görülmektedir (basık olan çizgiler).

Şekil 27

Test Bilgi ve Standart Hata Fonksiyonları



Bu grafikte ise mavi eğim test bilgi fonksiyonunu yansıtmaktadır. Test bilgi fonksiyonuna bakılarak x ekseni üzerindeki farklı yetenek düzeyleri için bu testteki tüm maddeler kullanıldığında ulaşılacak bilgi miktarının ne olduğu belirlenebilmektedir. Görüldüğü üzere test en çok orta yetenek düzeyindeki bireyler için bilgi sağlarken uçlardaki bireyler (çok başarılı ya da çok başarısız) için daha az bilgi sunmaktadır. Pembe eğri ise standart hatayı temsil etmektedir. Grafikte görüldüğü gibi standart hata, bilgi ile ters orantılıdır. Bu grafiklerin ardından kestirilen madde parametreleri ve madde – model uyum istatistiklerinin yer aldığı bir tablo sunulmaktadır. Bu tabloya ait ekran görüntüsü Şekil 28'de verilmiştir.

Şekil 28

Madde Parametreleri

	a	SE(a)	b	SE(b)	c	SE(c)	d	SE(d)	SX2-value	df	p-value
Item1	0.70	0.08	-0.16	0.07	0.00	NA	1.00	NA	6.54	13.00	0.92
Item2	0.65	0.07	-0.47	0.09	0.00	NA	1.00	NA	16.96	13.00	0.20
Item3	0.42	0.07	-2.27	0.38	0.00	NA	1.00	NA	29.22	14.00	0.01
Item4	0.38	0.08	-3.49	0.68	0.00	NA	1.00	NA	11.37	14.00	0.66
Item5	0.82	0.10	-2.04	0.21	0.00	NA	1.00	NA	14.79	12.00	0.25

Görüldüğü gibi Şekil 28; *a*, *b*, *c*, *d* parametreleri ve bu parametrelere ait standart hata değerleri ile S_χ^2 istatistiklerini içermektedir. 2PLM ile ölçekleme yapıldığından tüm maddeler için *c* parametresi 0 ve *d* parametresi 1 olarak atanmıştır. Öte yandan madde - model uyumunu gösteren S_χ^2 istatistiklerine ait *p* değerleri incelenip model ile uyum göstermeyen maddeler tespit edilmelidir. Bunun için ölçüt kabul edilecek alfa düzeyine test geliştiricisi karar verir. Bu uygulamada I. tip hatayı azaltmak amacıyla alfa düzeyi 0,01 olarak seçilmiştir. Buna göre tüm maddelerin model - veri uyumunu sağladığı anlaşılmıştır.



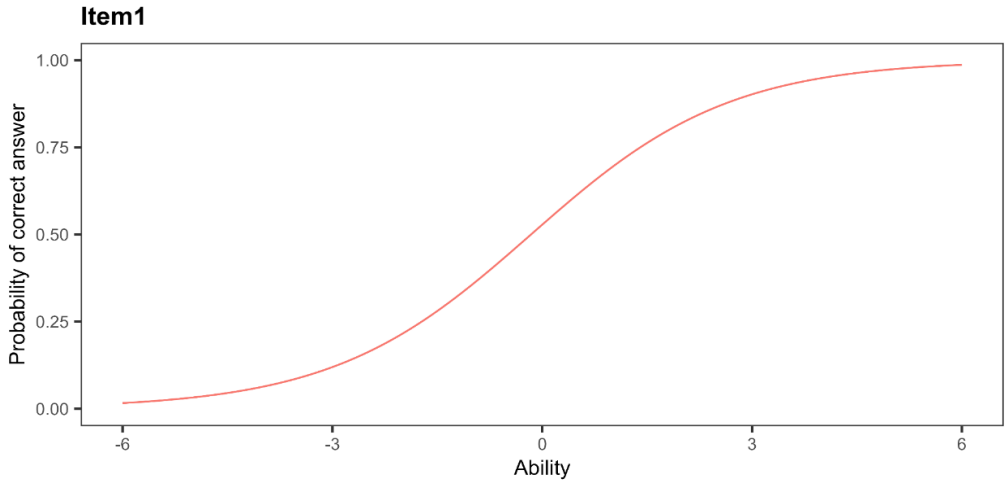
Eğer alfa düzeyi 0,05 seçilirse, 3 numaralı maddenin testten çıkarılması gerekir.

Madde parametrelerinin ardından her bir kişiye ait toplam ham puan, *z* ve T-puanları ile MTK'ye göre yetenek kestirimlerinin yer aldığı bir tablo özeti sunulmaktadır. Alttaki Download abilities (yetenekleri indir) butonu ile testi alan tüm bireylere ait puanlar indirilebilmektedir. Son olarak standartlaştırılmış *z*-puanları ile MTK yetenek kestirimleri arasındaki korelasyon katsayısı sunulmaktadır. Örnek uygulama için *z*-puanları ile teta yetenek kestirimleri arasındaki Pearson momentler çarpım korelasyon katsayısı 0,982 olarak raporlanmıştır.

Item (madde) alt sekmesinde ise her bir madde için ayrı ayrı madde karakteristik eğrileri, madde bilgi fonksiyonları ve madde parametreleri verilmektedir. Örnek olarak Madde 1 için madde karakteristik eğrisi ve madde bilgi fonksiyonu sırasıyla Şekil 29 ve Şekil 30'da sunulmuştur.

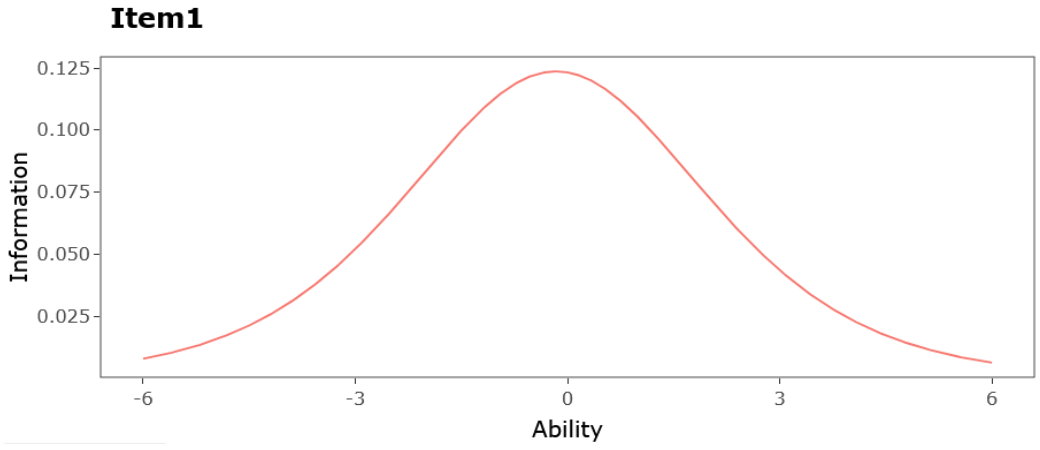
Şekil 29

Madde 1 için Madde Karakteristik Eğrisi



Şekil 30

Madde 1 için Madde Bilgi Fonksiyonu



Görüldüğü üzere bu defa tek bir madde için madde karakteristik eğrisi ve madde bilgi fonksiyonu sunulmaktadır. Özellikle madde havuzunda çok sayıda madde olduğunda tüm maddeleri aynı grafik üzerinde incelemek güçleşir. Bu durumda her maddeyi ayrı ayrı incelemek test geliştiricisine kolaylık sağlar.

5. SONUÇ

Bu bölüm ölçme sonuçları üzerinde yapılan istatistiksel işlemlerden test ve madde istatistiklerine odaklanmış hem KTK hem de MTK bağlamında yapılabilecek analizler Shiny Item Analysis ile örnek bir uygulama üzerinden gösterilmiştir. Test istatistikleri başlığı altında ele alınan merkezî eğilim ve yayılım ölçüleri birer betimsel istatistik olup ölçme işleminden elde edilen puanlar ile yapılabilecek hipotez testleri (T-testleri, varyans analizleri, regresyon, vb.) bölümün kapsamı dışında tutulmuştur. Dileyen okuyucular temel istatistik üzerine yazılmış kaynaklar aracılığıyla bu analizlere ilişkin okuma yapabilirler.

KAYNAKÇA | REFERENCES

Arıcı, H. (2006). *İstatistik: Yöntemler ve uygulamalar*. Meteksan.

Arıkan, S. & Aybek, E. C. (2022). A special case of Brennan's index for tests that aim to select a limited number of students: A Monte Carlo simulation study. *Educational Measurement: Issues and Practice, 41*, 35-49. https://doi.org/10.1111/emip.12528

Aybek, E. C. (2021). catIRT tools: A "Shiny" application for item response theory calibration and computerized adaptive testing simulation. *Journal of Applied Testing Technology, 22*(1). 12-24.

Baker, F. (2001). *The basics of item response theory*. ERIC Clearinghouse.

Baker, F. B. & Kim, S. H. (2017). *The basics of item response theory using R*. Springer.

Brennan, R. L. (1972). A generalized upper-lower item discrimination index. *Educational and Psychological Measurement, 32*(2), 289–303. https://doi.org/10.1177/001316447203200206

Crocker, L. & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.

Erkuş, A. (2003). *Psikometri üzerine yazılar*. Türk Psikologlar Derneği.

Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme - I: Temel kavramlar ve işlemler*. Pegem.

Evans, C. H. & Ildstad, S. T. (2001). *Small clinical trials: Issues and challenges*. National Academy.

Gronlund, N. E. (1977). *Constructing achievement tests*. Prentice-Hall.

Gulliksen, H. (1950). *Theory of mental tests*. John Wiley and Sons.

Haladyna, T. M. & Downing, S. M. (1993). How many options is enough for a multiple-choice test item? *Educational and Psychological Measurement, 53,* 999-1010. https://doi.org/10.1177/0013164493053004013

Hambleton, R. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Springer.

Kaptan, S. (1998). *Bilimsel araştırma ve istatistik teknikleri*. Tekışık Ofset.

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*(1), 17-24. https://doi.org/10.1037/h0057123

Köklü, N., Büyüköztürk, Ş. & Çokluk-Bökeoğlu, Ö. (2007). *Sosyal bilimler için istatistik*. Pegem.

Lord, F. M. (1952). The relation of the reliability of multiple-choice tests to the distribution of item difficulties. *Psychometrika, 17*(2), 181-194. https://doi.org/10.1007/BF02288781

Martinkova, P. & Hladka, A. (2023). *Computational aspects of psychometric methods*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003054313

Martinkova, P. & Drabinova, A. (2018). ShinyItemAnalysis for teaching psychometrics and to enforce routine analysis of educational tests. *The R Journal, 10*(2), 503-515. https://doi.org/10.32614/RJ-2018-074

Metsamuuronen, J. (2020). Somers' D as an alternative for the item–test and item-rest correlation coefficients in the educational measurement settings. *International Journal of Educational Methodology, 6*(1), 207–221. https://doi.org/10.12973/ijem.6.1.207

Murphy, K. R. & Davidshofer, C. O. (2014). *Psychological testing: Principles and applications* (6. baskı). Pearson.

Popham, W. J. (2014). *Classroom assessment: What teachers need to know*. Pearson.

Thompson, N. (2020). *What is classical item difficulty (p value)?. https://assess.com/classical-item-difficulty-p-value/*

Turgut, M. F. (1997). *Eğitimde ölçme ve değerlendirme metotları.* Yargıcı.

Turgut, M. F. & Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Pegem.

Woolley, T. W. (1998). A glossary of common statistical terms. *Journal of Pharmacy Practice, 11*(3). 219-225. https://doi.org/10.1177/089719009801100311

YAZAR HAKKINDA | ABOUT AUTHOR

DOÇ. DR. EREN CAN AYBEK

Pamukkale Üniversitesi | eaybek[at]pau.edu.tr

ORCID: 0000-0003-3040-2337

Eren Can Aybek lisans derecesini 2009 yılında Balıkesir Üniversitesi Fizik Öğretmenliği bölümünden almıştır. Yüksek lisans ve doktora eğitimini ise Ankara Üniversitesi Ölçme ve Değerlendirme bölümünde tamamlamıştır. 2017 yılından beri Pamukkale Üniversitesi Eğitim Fakültesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalında öğretim üyesi olarak görevini sürdürmektedir. Aynı zamanda 2021 yılında Boğaziçi Üniversitesi Bireye Uyarlanmış Test Laboratuvarında araştırmacıdır. Çalışma alanları çoğunlukla Bilgisayar Ortamında Bireye Uyarlanmış Testler (BOBUT) ve Madde Tepki Kuramını (MTK) içermektedir. Ayrıca otomatik puanlama, BOBUT simülasyonu ve faktör analizine yönelik çeşitli psikometrik yazılımlar geliştirmiştir.

ASSOC. PROF. DR. EREN CAN AYBEK

Pamukkale University | eaybek[at]pau.edu.tr

ORCID: 0000-0003-3040-2337

Eren Can Aybek received his bachelor's degree in Physics Education from Balıkesir University in 2009. He completed both his master's and doctoral studies in Measurement and Evaluation at Ankara University. Since 2017, he has been serving as a faculty member in the Department of Educational Measurement and Evaluation at the Faculty of Education, Pamukkale University. Since 2021, he has also been a researcher at the Adaptive Testing Laboratory at Boğaziçi University. His research interests primarily include Computerized Adaptive Testing (CAT) and Item Response Theory (IRT). He has also developed several psychometric software applications for automated scoring, CAT simulation, and factor analysis.