Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi

Journal of Measurement and Evaluation in Education and Psychology

ISSN: 1309 - 6575

Genellenebilirlik Kuramı Karar Çalışmalarında Kullanılan Farklı Varyans Bileşenleri Kestirim Yöntemlerinin Karşılaştırılması*

Comparing Different Variance Component Estimation Methods Used in Generalizability Theory Decision Studies

Eren Halil ÖZBERK** Selahattin GELBAL***

Öz

Bu araştırmanın amacı; verilerin normal dağılım varsayımına sahip olmadığı durumlarda farklı varyans bileşenleri kestirme yöntemlerini, genellenebilirlik ve Phi katsayılarının yanında kullanılması önerilen (Brennan, 2001; Kane, 1999) evren puanı-hata oranı ve hata-tolerans indisleri yardımı ile karşılaştırmaktır. Araştırma; iki yüzeyli bir verinin normal dağılım varsayımına sahip olmadığı durumda varyans bileşenlerini belirlemede ANOVA yöntemi ile bootstrap yöntemlerini farklı katsayı ve indisler yardımı ile karşılaştırmaktadır. Araştırmada $b\,x\,m$ desenine uygun ve birey-madde matrisi oluşturacak şekilde tek faktörlü olarak 60 x 5 şeklinde normal dağılıma sahip olmayan iki kategorili puanlanan veri seti üretilmiş, elde edilen veriler 25 replikasyon sonucu nihai halini almıştır. Örnekleme simülasyonu aşamasında ise verilerin simülasyonundan elde edilen veriler, desenine uygun olarak yüzeylere göre 1000 kere yeniden örneklenmiştir (bootstrap). Tüm yüzeylere göre ANOVA ve bootstrap yöntemleri kullanılarak standart hatalar, varyans bileşenleri, mutlak ve bağıl hatalar kestirilmiştir.Araştırma sonuçlarına göre normal dağılım göstermeyen ve iki kategorili puanlanan veriler üzerinde hesaplanan evren puanı-hata değeri en iyi $boot-b$ prosedüründe kestirilirken, hata tolerans değeri en iyi $boot-m$ prosedüründe kestirilmiştir. Bu bakımdan $boot-m$ prosedürünün daha geçerli bilgiler verdiği, $boot-b$ prosedürünün de G Kuramı çalışmalarında evren puanlarını belirlemede daha kesin kestirimler yaptığı sonucuna varılmıştır.

Anahtar Kelimeler: Genellenebilirlik kuramı, varyans bileşenleri kestirimi, yeniden örnekleme

Abstract

The aim of this research is to compare various variance component estimations procedures with using signal noise ratio and error tolerance ratio which is offered with generalizabiity and Phi coefficients in non-normal distrubitions (Brennan, 2001; Kane, 1999). This research compares variance components estimations with using ANOVA and bootstrap procedures in non-normal disturbitions in one facet design G studies. Data were gathered with using two seperate procedures (a) data simulation and (b) sampling simulation. In data simulation part, it's been simulated a non-normal dichotomous data set which fits to unidimensional person-item matrix 60x5 which fits to $b\,x\,m$ design. All the simulations replicated 25 times. In sampling simulation sections datas, gathered from data simulation sections has been bootstrapped 1000 times according to the each facet. Standart errors, variance components, relative and absolute error are estimated according to the each facets with using ANOVA and bootstrap procedures. The results also show that in non-normal dichotomously scored datas best signal-noise ratio has estimated in $boot-b$ procedure, and best error-tolerance ratio has been estimated in $boot-m$ procedure. Thus, $boot-m$ procedures gives more valid estimations and $boot-b$ procedure gives more reliable and precise estimations of universe scores in G studies rather than other procedures.

Key Words: Generalizability theory, variance component estimation, bootstrap

* Yazarın yüksek lisans tezinin bir kısmını içeren bu çalışma, The 78th Annual Meeting of the Psychometric Society Kongresi'nde sözlü bildiri olarak sunulmuştur.

** Eğitim ve Öğretim Planlamacısı, Hacettepe Üniversitesi Rektörlüğü, Ankara - TÜRKİYE, erenozberk@gmail.com

*** Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara - TÜRKİYE, sgelbal@gmail.com

GİRİŞ

Genellenebilirlik kuramı (G Kuramı), klasik test kuramındaki güvenirlik hesaplamalarında kullanılan gerçek puanlar ve gözlenen puanların farkından oluşan hatanın hesaplanmasında var olan sınırlılıklara tepki olarak Cronbach ve arkadaşları (1963) tarafından ortaya atılmıştır. Genellenebilirlik kuramının amacı, ölçme sonuçlarını farklı varyans kaynaklarına ayırarak, yorumlayarak ve tanımlayarak, ölçme konusu olan bireyler ya da objelerin gözlenen puanlarının evren puanlarına doğrulukla genellenmesini sağlamaktır. Kuram aynı zamanda, davranış ölçümlerinde güvenirliği değerlendirirken, güvenilir gözlemlerin tasarımını, araştırılmasını ve kavramsallaştırılmasını sağlayan istatistiksel bir kuramdır (Cronbach, Gleser, Nanda ve Rajaratnam, 1972; Shavelson ve Webb, 1991; Brennan, 2001). G kuramı, ölçme işlemlerinin tekrar edilmesi ile elde edilecek olan gözlenen puanlar arasındaki tutarsızlıkların, kaynağının ve miktarının belirlenmesine de olanak vermektedir.

G kuramı; madde, zaman, puanlayıcı ve benzeri hata kaynaklarını, yüzey (facet) olarak adlandırır. Yüzey, deneysel desenlerin literatüründeki "faktör" kavramına benzer. Bu yüzeylerin düzeyleri (levels) (madde sayısı, puanlayıcı sayısı), koşullar (conditions) olarak adlandırılır. G kuramında kullanılan yüzey ve koşul ifadeleri geleneksel varyans analizinde faktör ve düzey kavramlarına karşılık gelir. (Crocker ve Algina, 1986; Shavelson ve Webb, 1991; Brennan, 2001). Bu bakımdan kuramın temeli varyans analizi (ANOVA) üzerine kurulmuştur. Varyans analiziyle toplam varyans desendeki bağımsız değişkenlere bölünür. Böylece ölçme sonuçları farklı varyans kaynaklarına ayrılarak bireylerin ya da objelerin gözlenen puanlarının evren puanlarına (gerçek puanlarına) genellenebilmesi sağlanır (Shavelson ve Webb, 1991).

G kuramı, yüzey sayısına göre tek veya daha fazla desenin oluşturulması ile güvenirlik ve Phi katsayılarını hesaplayabilmektedir. Bu araştırmada tek yüzeyli çaprazlanmış desen kullanılmıştır.

Genellenebilirlik Çalışması

G çalışması farklı koşullardaki ölçmelerden kaynaklanan hataları araştırmak ve yüzeylere ait varyans bileşenlerini hesaplamak için tasarlanır. Bu nedenle bir G çalışmasının amacı farklı yüzeylerdeki gözlenen puanların varyansını tespit etmek ve bu varyans bileşenlerini ayrıştırmaktır (Rentz, 1987). Diğer bir ifadeyle G çalışmaları mümkün olduğunca en kapsamlı şekilde, kabul edilebilir gözlemler evrenini tanımlamalıdır (Shavelson ve Webb, 1991). G çalışmalarının amacı evrende değişkenliğin olası kaynakları hakkında bilgi sağlamaktır. G çalışmaları mümkün olduğunca olası yüzeyleri desenlerine dâhil etmeyi ve belirlemeyi sağlamalıdır. G çalışmalarının özünde her bir bireye ait ortalama puanların hesaplanması yatar. Ancak ortalama puanlar gözlenebilir değerler değillerdir. Bu yüzden, birey-madde puanlarının gözlenmesi ile her bir birey için ortalama puan hesaplanmış olur. Bu hesaplamalar Eşitlik 1'de verilmiştir.

$$
\begin{aligned}
X_{bm} &= \mu && \text{(Genel Ortalama)} \\
&+ \mu_b - \mu && \text{(Birey etkisi - } v_b\text{)} \\
&+ \mu_m - \mu && \text{(Madde etkisi - } v_m\text{)} \\
&X_{bm} - \mu_b - \mu_m - \mu && \text{(Artık etki - } v_{bm}\text{)}
\end{aligned} \tag{1}
$$

Böylece her bir bireye ait puan değeri Eşitlik 2'deki denklem ile hesaplanmaktadır.

$$X_{bm} = \mu + v_b + v_m + v_{bm} \tag{2}$$

Kullanılan her denklem (μ hariç) rastgele etki olarak adlandırılır. Çünkü yapılan işlem popülasyondan ve evrenden rastgele örnekleme sürecidir ve bu işlemlerin birbirinden bağımsız olduğu varsayılır.

Genellenebilirlik Çalışmalarında Varyans Bileşenlerini Belirleme

Her bir puan etkisine bağlı olarak varyans bileşenlerini kestirme G kuramı çalışmalarında çok önemlidir (Brennan, 2003). Varyans bileşenlerini kestirmede birçok yol kullanılır ancak bunlardan en yaygın kullanılanı, ortalama puanlara dayalı kestirim yapan ANOVA yöntemidir.

ANOVA yöntemi kullanılarak varyans bileşenlerini belirlemede ortalama kareler değeri kullanılır. Her bir yüzey için beklenen ortalama kareler değerleri Eşitlik 3'te verilmiştir.

$$\begin{aligned} EMS(b) &= \sigma^2(bm) + n_m\sigma^2(b) \\ EMS(m) &= \sigma^2(bm) + n_b\sigma^2(m) \\ EMS(bm) &= \sigma^2(bm) \end{aligned} \tag{3}$$

Yüzeylere ait varyans bileşenleri değerleri ise Eşitlik 4'te verilmiştir.

$$\begin{aligned} \sigma^2(b) &= [MS(b) - MS(bm)]/n_m \\ \sigma^2(m) &= [MS(m) - MS(bm)]/n_b \\ \sigma^2(bm) &= MS(bm)] \end{aligned} \tag{4}$$

Varyans bileşenlerinin farklı yollardan kestirimlerine ilişkin ilk çalışmalar Searle (1987) tarafından uygulanan maksimum olabilirlik (ML), kısıtlı maksimum olabilirlik (REML) ve minimum ikinci derece yansız kesitirim (MINQUE) yöntemleri ile ortaya çıkmıştır. Daha sonra standart hatalara dayalı olarak kestirilen varyans bileşenleri yöntemlerinden olan bootstrap (yeniden örnekleme) yöntemi de araştırmacılar tarafından ortaya atılmıştır (Brennan, Harris ve Hanson, 1987; Luecht ve Smith, 1989; Shavelson ve Webb, 2004). Dağılımın normal olmadığı durumlarda, örneğin 1'den 5'e kadar puanlanmış bir ölçekte normalliğin sağlanmasının bir sorun olacağı ve normallik varsayımını gerçekleştirmeyen yeni bir varyans bileşeni kestirme yönteminin gerekliliği belirtilmiştir (Brennan, 2001). Bu sayede bootstrap yöntemi ile veri setinden şansa bağlı örnekler alınarak, normallik varsayımına bakılmaksızın varyans bileşenlerinin kestirilmesinde kullanılmaya başlanmıştır (Moore, 2010; Leucht ve Smith, 1989; Brennan, 1987).

Bootstrap yöntemi özünde bir veri setinden yer değiştirme ile N bireylik B adet bootstrap örneği oluşturur. Daha sonra her bir bootstrap örneklemi için bir $\hat{\theta}_b = S(x_b)$ değeri kestirilir. Ve bu $\hat{\theta}_b$ değerinden yararlanarak standart hata değeri kestirilir. Bootstrap prosedürlerinin $b\,x\,m$ desenlerine uygulanması ise bazı kurallar çerçevesinde yapılmıştır. Bootstrap prosedürünün G kuramına uygulanmasında ilk olarak gözlenen puanlardan oluşan $n_b\ x\ n_m$ matrisinin oluşturması gerekir. Daha sonra yüzeylerden birinin sabit tutulup diğerinin n sayısı kadar yeniden örneklemeye sokulması ile değerler kestirilip, varyans bileşenleri ANOVA yöntemi tekrar kullanılarak ayrı ayrı kestirilir.

G kuramında, uygulanacak desenlere göre ($b x m$, $b x m x p$, $b x (m : p) \ldots$) varyans bileşenleri ve bileşenlere ait standart hataları kestirmek için çok farklı bootstrap prosedürleri kullanılabilir. Kullanılacak bu prosedürler, yüzeyin yeniden örneklenmesi isteğine göre de farklılık gösterir (Moore, 2010). Farklı bootstrap yöntemleri, farklı varyans bileşenleri kestirim değerleri sağladığı için en doğru ve kesin bir bootstrap prosedüründe bahsedilemez (Brennan, 2007; Brennan, Harris ve Hanson, 1987). Örnek olarak bir grup öğrencinin (b) birçok maddeden oluşan (m) testi yanıtladığını varsayalım. Bu durumda kullanılacak bootstrap prosedürü öğrencilerin yeniden örneklemeye katılıp maddelerin katılmadığı

($boot-b$); maddelerin yeniden örneklemeye katılıp öğrencilerin katılmadığı ($boot-m$) ve hem maddelerin hem de öğrencilerin yeniden örneklemeye katıldığı ($boot-bm$) olmak üzere üç durumda incelenebilir.

Karar Çalışması

K çalışması, puanların değişkenliğini belirlemede temel bir çerçeve çizmeye olanak tanır (Meyer, 2010; Brennan, 2003). Genellenebilirlik evreni yardımı ile her bir yüzeye ait gözlemlerin sayısı, ölçme prosedürleri ile işleme konularak karar çalışmaları adına örnekler oluşturulur.

Karar çalışmasında genellenebilirlik katsayısı (G) ve Phi (Φ) katsayısı olmak üzere iki tür güvenirlik katsayısı hesaplanabilir. Fakat hesaplanan bu güvenirlik katsayıları hata varyansına göre değişiklik göstermektedir. Bu hata varyansları mutlak (absolute) hata ve bağıl (relative) hata varyansı olarak adlandırılmaktadır. Mutlak hata; bir kişinin gözlenen puanı ve onun evren puanı arasındaki fark olarak tanımlanır. Bağıl hata ise; gözlenen sapma puanı ve evren sapma puanı arasındaki farklılıktır. Bağıl hata varyansı $\sigma^2(\delta)$ şekilde gösterilir ve klasik test kuramındaki hata varyansına karşılık gelir (Brennan, 2001).

Genellenebilirlik Katsayısı:

Genellenebilirlik katsayısı Cronbach ve arkadaşları tarafından (1972) tarafından ortaya konulmuştur. Genellenebilirlik katsayısı, bağıl kararlar vermek üzere kullanılan güvenirlik olarak açıklanmaktadır (Meyer, 2010). G katsayısı, evren puan varyansının beklenen gözlenen puan varyansına oranı olarak ifade edilir. G katsayının formülü Eşitlik 5'te verilmiştir.

$$E\rho^2 = \frac{\sigma^2(b)}{\sigma^2(b)+\sigma^2(\delta)} \tag{5}$$

Eşitlik 5'te de görüldüğü gibi $E\rho^2$ bağıl hata varyansını ($\sigma^2(\delta)$) kullanır.

Phi Katsayısı:

Phi (Φ) katsayısı; ilk olarak Brennan ve Kane (1977) kararlılık indeksi olarak tanımlanmış ve mutlak kararlar için puanların güvenirliği şeklinde açıklanmıştır. K katsayısı, mutlak hata varyansını ($\sigma^2(\Delta)$) içeren bir katsayıdır. Phi katsayının formülü Eşitlik 6'da verilmiştir.

$$\Phi = \frac{\sigma^2(b)}{\sigma^2(b)+\sigma^2(\Delta)} \tag{6}$$

Hata-Tolerans Oranı (Error-Tolerance Ratio) (E/T)

Hata-tolerans oranının ortaya çıkmasının başlıca nedeni belirli koşullar altında yapılan ölçmelerin ne kadar kesin olduğunun sorgulanmasıdır. Hataların ciddi problemlere yol açmadan önce; ne kadar hatanın belirli durumlar için tolere edilebilir olması gerektiğinin önemini vurgulayan Kane (1996), ölçmenin duyarlılığını nicelemek adına "hata-tolerans oranını" ortaya atmıştır. Ayrıca Kane (1996) toleransı konsept olarak geçerliğe daha yakın bulduğunu, çünkü hatalardan elde edilen toleransın kararların yorumlanmasındaki ölçme hatalarının etkisinde olduğunu belirtmiştir. Kane (1996) hatalara ait toleranslarını değerlendirmek için ilk olarak hata-tolerans oranını tanıtmıştır. Gerçek puanlar modeline dayanarak E/T; "hatanın standart sapmasının, gerçek puanların standart sapmasına oranı" şeklinde tanımlamıştır ve formülü Eşitlik 7'de verilmiştir:

$$E/T = \frac{\sigma_E}{\sigma_T} \tag{7}$$

E/T eşitliğinin özünde yatan ifade; testi alan her bir birey için hata toleransları bireylerin gerçek sapma puanları şeklinde tanımlanabilir.

Evren Puanı-Hata Oranı (Signal-Noise Ratio) (S/N ***)***

Hata varyanslarından anlam çıkarmada, hata varyansının büyüklüğünün evren puan varyansları ile oranı daima yorumlama açısından önemli bir durum teşkil eder. Bu durumu ortaya çıkarıcı işlemlerden biri evrendeki puan varyanslarının hata varyansına bölümü ile bulunabilir. Bu oran evren puanı-hata oranı olarak adlandırılır. E/T değerinin tersi olarak bilinir. Formülü eşitlik 8 de verilmiştir:

$$S/N = \frac{1}{E/T} = \frac{\sigma_T}{\sigma_E} \qquad (8)$$

Evren puanı-hata oranında, evrende sınavı alan bireyler arasındaki fark "sinyal (signal)" olarak ele alınır ve gerçek puanların standart sapması bu sinyalin bütün bir gücü şeklinde bir tanımlama yapılmıştır. Aynı popülasyon üzerinden elde edilen hatalar ise "gürültü (noise)" olarak adlandırılır ve standart hata da sinyali engelleyen bir indeksmiş gibi düşünülebilir (Kane, 1996).

Güvenirlik, hem E/T hem de S/N oranları şeklinde ifade de edilebilir:

$$\rho_{XX} = \frac{1}{1+(E/T)^2} = \frac{(S/N)^2}{(S/N)^2+1} \qquad (9)$$

G kuramında en çok kullanılan ve bulgular kısmında rapor edilen istatistikî değer bağıl hataya ($\sigma^2(\delta)$) bağlı genellenebilirlik katsayısı ve mutlak hataya bağlı ($\sigma^2(\Delta)$) Phi katsayısıdır. Kullanılan bu katsayıların yanında yeni olarak; kullanılması Brennan (2003) tarafından önerilen Hata-Tolerans Oranı indisi ve Evren Puanı-Hata Oranı indisleri hata varyanslarının büyüklüğünün evren puan varyansları ile karşılaştırmasına olanak vermesi ve belli koşullar altında elde edilen ölçme sonuçlarından ortaya çıkan hataların, evren puanlarını belirlemede ne kadarının tolere edilebileceğini ortaya koyması açısından daha bilgi verici olduğunu vurgulamıştır. Kane (1996), hatalar için elde edilen tolerans değerinin ölçme sonuçlarının kullanılış amaçlarına göre farklılık göstereceğini belirtmiştir. Belirli durumlarda ölçme sonuçlarından beklenen çıktıların daha kesin bir şekilde elde edilmesi adına hataların en aza indirilmesi gerektiğini vurgulamıştır. Bu duruma örnek olarak ise laboratuar dışında elde edilen bir ağırlık (gram) ölçümünün standart hatasına bakılmasının gereksiz olabileceğini ama aynı ölçümün laboratuar ortamında yapılması halinde hata oranının önem kazanacağını söylemiştir. Bu bakımdan hangi koşullarda ne kadar hatanın tolere edileceği önem kazanır. Genellenebilirlik kuramında, az sayıda örneklem ve normal dağılım varsayımının karşılanmadığı durumlarda hangi varyans bileşeni kestirim yönteminin daha geçerli sonuçlar vereceği farklılık gösterebilmektedir. Elde edilen kestirimlerin daha farklı indisler yardımı ile yorumlanmasıyla sonuçların geçerliğinin artacağı düşünülmektedir.

Araştırmanın Amacı

Bu araştırmanın amacı; verilerin normal dağılım varsayımına sahip olmadığı durumlarda farklı varyans bileşenleri kestirme yöntemlerini, genellenebilirlik ve Phi katsayılarının yanında kullanılması önerilen (Brennan, 2001; Kane, 1999) evren puanı-hata oranı (S/N) ve hata-tolerans oranı (E/T) indisleri yardımı ile karşılaştırmaktır.

Çalışmada ilk olarak ANOVA yöntemi kullanılarak varyans bileşenleri kestirilip, G kuramı K çalışmalarında kullanılan katsayı ve indisler yardımı ile yorumlanacaktır.

İkinci olarak varyans bileşenleri, yeniden örnekleme tekniği olarak bilinen ve son zamanlarda G kuramı çalışmalarında araştırmacılar tarafından kullanılan (Wiley, 2001; Brennan, 2003; Moore, 2010) bootstrap prosedürleri yardımı ile farklı yüzeyler göz önüne alınarak belirlenecek; aynı şekilde G kuramı K çalışmalarında kullanılan katsayı ve indisler yardımı ile yorumlanacaktır.

YÖNTEM

Araştırma; verilerin normal dağılım varsayımına sahip olmadığı durumlarda varyans bileşenlerini belirlemede kullanılan ANOVA yöntemi ile bootstrap yöntemlerini farklı katsayı ve indisler yardımı karşılaştırmayı amaçlamaktadır. Araştırma yöntem karşılaştırması yapılması bakımından kuramsal bir araştırmadır.

Veri Toplama Teknikleri

Bu çalışmada verilerin elde edilmesi kısmı iki başlık altında toplanacaktır:

1. Veri simülasyonu
2. Örnekleme simülasyonu

İlk kısımda genellenebilirlik kuramı çalışmalarına uygun bir veri seti, R paket programı yardımı ile elde edilecektir. İkinci kısımda elde edilen veriler kullanılarak, S PLUS programı yardımı ile bootstrap prosedürleri uygulanacak ve yüzeylere göre varyans bileşenleri kestirilecektir.

Veri Simülasyonu

İlk aşamada $b\,x\,m$ desenine uygun ve birey-madde matrisi oluşturacak şekilde tek faktörlü olarak 60 x 5 şeklinde normal dağılıma sahip olmayan iki kategorili puanlanan veri seti üretilmiştir. Veri simülasyonu için R yazılımında bulunan {psych} paketindeki "*sim.VSS*" komutu kullanılmıştır, elde edilen veriler 25 replikasyon sonucu nihai halini almıştır.

Örnekleme Simülasyonu

İkinci aşamada verilerin simülasyonundan elde edilen veriler, $b\,x\,m$ desenine uygun olarak her bir yüzeye göre 1000 kere yeniden örneklenmiştir (bootstrap).

İşlem

Bootstrap yöntemi özünde bir veri setinden yer değiştirme ile N bireylik B adet bootstrap örneği oluşturur. Daha sonra her bir bootstrap örneklemi için bir $\hat{\theta}_b = S(x_b)$ değeri kestirilir. Ve bu $\hat{\theta}_b$ değerinden yararlanarak standart hata değeri kestirilir. Bootstrap prosedürlerinin $b\,x\,m$ desenlerine uygulanması ise bazı kurallar çerçevesinde yapılmıştır. Bootstrap prosedürünün G kuramına uygulanmasında ilk olarak gözlenen puanlardan oluşan $n_b\;x\;n_m$ matrisinin oluşturması gerekir. Daha sonra yüzeylerden birinin sabit tutulup diğerinin n sayısı kadar yeniden örneklemeye sokulması ile değerler kestirilip, varyans bileşenleri ANOVA yöntemi tekrar kullanılarak ayrı ayrı kestirilir.

Yüzeylere göre varyans bileşenlerinin ve standart hataların belirlenmesinde kullanılan katsayı ve formüller aşağıdaki gibidir (Wiley,2001):

$\underline{boot-b}$

$$\hat{\sigma}^2(b) = \frac{n_b}{n_b - 1}\hat{\sigma}^2(b \mid boot-b)$$

$$\hat{\sigma}^2(m) = \hat{\sigma}^2(m \mid boot-b) - \frac{1}{n_b - 1}\hat{\sigma}^2(bm \mid boot-b)$$

$$\hat{\sigma}^2(bm) = \frac{n_b}{n_b - 1}\hat{\sigma}^2(bm \mid boot-b)$$

$\underline{boot-m}$

$$\hat{\sigma}^2(m) = \frac{n_m}{n_m - 1}\hat{\sigma}^2(m \mid boot-m)$$

$$\hat{\sigma}^2(b) = \hat{\sigma}^2(b \mid boot-m) - \frac{1}{n_m - 1}\hat{\sigma}^2(bm \mid boot-m)$$

$$\hat{\sigma}^2(bm) = \frac{n_m}{n_m - 1}\hat{\sigma}^2(bm \mid boot-m)$$

$\underline{boot-bm}$

$$\hat{\sigma}^2(b) = \frac{n_b}{n_b - 1}\hat{\sigma}^2(b \mid boot-b,m) - \frac{n_b}{(n_b-1)(n_m-1)}\hat{\sigma}^2(bm \mid boot-b,m)$$

$$\hat{\sigma}^2(m) = \frac{n_m}{n_m - 1}\hat{\sigma}^2(m \mid boot-b,m) - \frac{n_m}{(n_b-1)(n_m-1)}\hat{\sigma}^2(bm \mid boot-b,m)$$

$$\hat{\sigma}^2(bm) = \frac{n_b n_m}{(n_b-1)(n_m-1)}\hat{\sigma}^2(bm \mid boot-m)$$

<u>*Standart hata*</u> (Efron ve Tibshirani, 1993):

$$SH_B(\widehat{\theta}) = \{\frac{1}{B-1}\sum_{b=1}^{B}(\widehat{\theta}^{*b} - \overline{\widehat{\theta}^{*}})^2\}^{1/2}$$

$$\overline{\widehat{\theta}^{*}} = \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^{*b}$$

Verilerin Analizi

Verilerin analizinde, elde edilen varyans bileşenleri yardımı ile G kuramı K çalışmalarında kullanılan G katsayısı, Φ katsayısı, S/N ve E/T indisleri belirlenmiştir. Verilerin analizinde ANOVA yöntemi için katsayı ve indisler EduG programı yardımı ile belirlenmiştir. Varyans bileşenlerini bootstrap yöntemi ile kestirmek için S-PLUS programı kullanılmıştır.

BULGULAR

Araştırmada ilk olarak tek yüzeyli desen için ANOVA ve bootstrap yöntemlerinin varyans bileşenleri kestirimleri karşılaştırılmıştır.

Araştırmada simülasyon sonucu ile elde edilen verilere ANOVA ve bootstrap prosedürleri uygulanmış ve her bir yüzeye ait varyans bileşenleri değerleri ile bağıl ve mutlak hata değerleri Tablo 1'de verilmiştir.

Tablo 1. *b x m Deseni için Varyans Bileşenleri Tahmini ($n_b = 60, n_m = 5$)*

	$\sigma^2(b)$	$\sigma^2(m)$	$\sigma^2(bm)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
ANOVA	0,0393	0,0136	0,0630	0,0126	0,0153
boot-b	0,0380	0,0140	0,0640	0,0128	0,0156
boot-m	0,0420	0,0130	0,0602	0,0120	0,0146
boot-bm	0,0403	0,0138	0,0605	0,0121	0,0149

Tablo 1 incelendiğinde, $\sigma^2(b)$ değeri için en yüksek değer, bireylerin sabit tutulup, yalnızca maddelerin yeniden örneklemeye alındığı $boot - m$ olduğu görülmüştür. $\sigma^2(m)$ değeri için en yüksek değer, maddelerin sabit tutulduğu, bireylerin yeniden örneklemeye alındığı $boot - b$ yöntemi olmuştur. Elde edilen sonuçlar geçmişte yapılan çalışmaları destekler nitelikte bulunmuştur (Moore, 2010; Wiley, 2001; Luecht ve Smith, 1989; Brennan, Harris ve Hanson, 1987).

Varyans bileşenlerini belirlemede standart hata kestirilmesi için ANOVA ve bootstrap yöntemleri her bir yüzeye göre uygulanıp Tablo 2'de verilmiştir.

Tablo 2. $b\ x\ m$ *Deseni Varyans Bileşenleri Tahmini için Standart Hata Kestirimleri* ($n_b = 60, n_m = 5$)

	SH(b)	*SH*(m)	*SH*(bm)	*SH*(δ)	*SH*(Δ)
ANOVA	0,0094	0,0084	0,0057	0,113	0,123
boot-b	0,0092	0,0087	0,0058	0,112	0,124
boot-m	0,0098	0,0080	0,0055	0,109	0,120
boot-bm	0,0095	0,0085	0,0055	0,110	0,121

Standart hataların değerleri incelendiğinde, ANOVA yönteminin bootstrap yöntemine göre standart hata değerlerini kestirmede net bir üstünlüğü bulunamamıştır. Farklı yüzeylere göre standart hata değerlerinin ANOVA yöntemine göre daha düşük ve daha yüksek olarak kestirildiği görülmektedir.

Çalışmada G ve Phi katsayılarını belirlemede kullanılacak en temel katsayılardan *SH*(δ) değeri ANOVA yöntemi ile (*SH*(δ)=0,113), *SH*(Δ) değerleri ise yeniden örnekleme tekniği olan bootstrap yöntemi ile (*SH*(Δ)=0,124) en iyi olarak kestirilmiştir.

ANOVA yöntemi ve bootstrap yöntemi kullanılarak elde edilen varyans değerlerinin, yüzeyleri açıklama yüzdeleri Tablo 3'te belirtilmiştir. ANOVA yöntemi referans olarak alındığında, bireylerin yeniden örneklemeye sokulduğu durumda ($boot - b$) bireylere ait varyans bileşeninin yüzdesi azalmakta iken maddelere ve ortak etkiye ait açıklama yüzdeleri artmaktadır.

Tablo 3. $b\ x\ m$ *Deseni için Varyans Bileşenleri Yüzdeleri* ($n_b = 60, n_m = 5$)

	b (%)	*m (%)*	*bm (%)*
ANOVA	33,91	11,73	54,36
boot-b	32,76	12,07	55,17
boot-m	36,46	11,28	52,26
boot-bm	35,17	12,04	52,79

Aynı şekilde maddelerin yeniden örneklemeye alındığı ($boot - m$) durumda birey varyanslarının açıklama yüzdeleri en yüksek oranda iken (%36,46), maddeler ve ortak etki azalmaktadır. Her iki yüzeyin yeniden örneklemeye sokulduğu durumda ise bireylere ve maddelere ait varyans bileşeni yüzdesi artarken, ortak etki azalmaktadır. Elde edilen bu değerler, G kuramında kullanılan katsayı ve indislerin de aynı yönde artıp azalacağının bir ön göstergesi olmaktadır (Moore, 2010).

Araştırmanın $b\,x\,m$ deseni için farklı varyans bileşeni kestirim yöntemleri uygulanmış, genellenebilirlik kuramı değerleri ve araştırmada önerilen indis değerleri Tablo 4'te karşılaştırmalı olarak verilmiştir.

Tablo 4. *$b\ x\ m$ Deseni için Farklı Varyans Bileşeni Kestirim Yöntemlerine Göre Elde Edilen G Kuramı Değerleri*

		ANOVA	$boot-b$	$boot-m$	$boot-bm$
Mutlak Hata	(Δ)	0,123	0,124	0,120	0,121
Bağıl Hata	(δ)	0,113	0,112	0,109	0,110
Mutlak Hata Varyansı	$\sigma^2(\Delta)$	0,015	0,016	0,015	0,015
Bağıl Hata Varyansı	$\sigma^2(\delta)$	0,013	0,013	0,012	0,012
Evren Puan Varyansı	$\sigma^2(\tau)$	0,039	0,038	0,042	0,040
Gözlenen Puan Varyansı	$ES^2(\tau)$	0,052	0,051	0,054	0,052
Genellenebilirlik Katsayısı	$E\rho^2$	0,757	0,748	0,777	0,769
Phi Katsayısı	Φ	0,720	0,709	0,742	0,731
Evren Puanı-Hata Oranı (δ)	S/N_δ	3,119	2,969	3,488	3,331
Evren Puanı-Hata Oranı (Δ)	S/N_Δ	2,565	2,436	2,869	2,712
Hata-Tolerans Oranı (δ)	E/T_δ	0,566	0,580	0,535	0,547
Hata-Tolerans Oranı (Δ)	E/T_Δ	0,624	0,641	0,590	0,607

Tablo 4.'te verilen $b\,x\,m$ desenine ANOVA yönteminde G katsayısı 0,757, Phi katsayısı da 0,720 olarak bulunmuştur. Yapılan G çalışması sonucu hata-tolerans indisi (E/T_δ)=0,566 ve (E/T_Δ)=0,624 olarak kestirilmiştir. Araştırmacılar bu değerin, tek bir duruma bağlı olarak değerlendirilme yapılması durumunda yüksek bir değer olduğunu belirtmişlerdir (Hagvet ve Hoglend, 2008; Cohen, Kane ve Kim, 2001). Böylesi bir değerlendirmede hata tolerans karesi değerinin 0,3 ün altında olması gerektiği, en iyi sonucun alınması ve yapılan ölçme sonuçlarının kesinliğini belirlemede ise bu değerin 0,2'nin altına çekilmesi gerektiği özellikle belirtilmiştir (Hagvet ve Hoglend, 2008). Hata tolerans indisine bakarak, var olan koşullar altında elde edilen ölçme sonuçlarından ortaya çıkan hatalar, evren puanlarını belirlemede tolere edilecek değer kabul edilebilir sınırlarda olmadığı yorumu yapılabilir.

Bu bakımdan ANOVA yöntemi ile kestirilen G ve Φ katsayısı her ne kadar güvenirlik anlamında kabul edilebilir bir değer aralığında olsa da ($\geq 0,7$), hata tolerans değerine bakıldığında bu ölçümün geçerli olmadığı görülmektedir.

Maddelerin sabit tutulup bireylerin yeniden örneklemeye alındığı $boot-b$ yönteminde G katsayısı 0.748, Phi katsayısı da 0.709 olarak bulunmuştur. Yapılan G çalışması sonucu hata-tolerans indisi (E/T_δ)=0.580 ve (E/T_Δ)=0.641 olarak kestirilmiştir. Hata tolerans değeri bu çalışmada da kabul edilen sınırlar içerisinde çıkmamıştır. Hata tolerans kareleri değerleri de 0,3'ün üzerinde olduğu için maddelerin sabit tutulup, bireylerin yeniden örneklemeye koyulduğu varyans bileşenleri kestirilmesi prosedürü sonucu ortaya çıkan hatalar, evren puanlarını belirlemede tolere edilecek değer kabul edilebilir sınırlarda olmadığı yorumu yapılabilir.

Bu bakımdan $boot-b$ yöntemi ile kestirilen G ve Φ katsayısı da her ne kadar güvenirlik anlamında kabul edilebilir bir değer aralığında olsa da, hata tolerans değeri bu ölçümün de geçerli olmadığını belirtmektedir.

Bireylerin sabit tutulup maddelerin yeniden örneklemeye alındığı $boot-m$ yöntemi ait G katsayısı 0.777, Phi katsayısı da 0.742 olarak bulunmuştur. Yapılan G çalışması sonucu hata-tolerans indisi (E/T_δ)=0.535 ve (E/T_Δ)=0.590 olarak kestirilmiştir. Hata tolerans değeri bu çalışmada diğer bootstrap yönteminin aksine kabul edilen sınırlar içerisinde çıkmıştır. Bağıl hataya bağlı hata tolerans kareleri değerleri de 0,3'ün altında olduğu için bireylerin sabit tutulup, maddelerin yeniden örneklemeye koyulduğu varyans bileşenleri kestirilmesi prosedürü sonucu ortaya çıkan hatalar, evren puanlarını belirlemede tolere edilecek değerin, kabul edilebilir sınırlar içinde olduğu yorumu yapılabilir. Bu bakımdan $boot-m$ yöntemi ile kestirilen G ve Φ katsayısı ölçümlerin kabul edilebilir bir değer aralığında olduğunu, hata tolerans indisi de çok yüksek olmayan G ve Φ katsayısına rağmen ölçümün geçerli olduğunu belirtmektedir.

Hem bireylerin hem de maddelerin yeniden örneklemeye alındığı $boot-bm$ yöntemine ait G katsayısı 0.769, Phi katsayısı da 0.731 olarak bulunmuştur. Yapılan G çalışması sonucu hata-tolerans indisi (E/T_δ)=0.547 ve (E/T_Δ)=0.607 olarak kestirilmiştir. Hata tolerans değeri bu yöntemde de kabul edilen sınırlar içerisinde çıkmıştır. Bağıl hataya bağlı hata tolerans kareleri değerleri de 0,3 olarak kestirildiğinden bireylerin ve maddelerin yeniden örneklemeye koyulduğu varyans bileşenleri kestirilmesi prosedürü sonucu ortaya çıkan hatalar, evren puanlarını belirlemede tolere edilecek değerin, kabul edilebilir sınırlar içinde olduğu yorumu yapılabilir. Bu bakımdan $boot-bm$ yöntemi ile kestirilen G ve Φ katsayısı ölçümleri de her ne kadar $boot-m$ değerlerinden düşük çıkmış olsalar da, bağıl hataya bağlı tolerans indisi, G ölçümün geçerli olduğunu belirtmektedir.

SONUÇLAR ve TARTIŞMA

Çalışmada tek yüzeyli desen için ANOVA ve bootstrap yöntemleri kullanarak kestirilen varyans bileşenleri değerleri karşılaştırılmıştır. Her bir yüzeye yönelik varyans bileşenleri değerleri incelendiğinde bireylere ait varyans bileşeni $\sigma^2(b)$ değerini kestirmede en iyi yöntemin bireylerin sabit tutulduğu, maddelerin yeniden örneklemeye alındığı $boot-m$ yöntemi olduğu görülmüştür. Maddelere ait varyans bileşenlerini kestirmede ise $\sigma^2(m)$ maddelerin sabit tutulduğu, bireylerin yeniden örneklemeye katıldığı $boot-b$ yöntemi en iyi sonuçları vermiştir. Araştırmada her bir yüzeye ait standart hata değerleri de kestirilmiştir.

Bağıl ve mutlak standart hataların en düşük değeri $boot-m$ için kestirilirken (*SH*(δ)=0,109), en yüksek değerler de ANOVA yöntemi için kestirilmiştir (*SH*(δ)=0,113). Moore (2010) en düşük değerin elde edilmesinin sebebinin yüzeye ait örneklem büyüklüğünün az olduğu durumda oluşacağını belirtmiştir. Bu bakımdan mutlak ve bağıl hata değerlerinin $boot-m$ prosedürü için düşük çıkması örneklem sayısının az olmasından ($n_m=5$) kaynaklandığı sonucuna ulaşılabilir. Otham (1995), standart hata kestiriminde ikili veriler ile yaptığı çalışmada ANOVA yöntemini daha uygun bulmuştur. Çalışmada elde edilen sonuçlar Otham (1985)'ın sonuçlarını doğrulamamaktadır. Bunun nedeni olarak bu çalışmada Wiley (2001) tarafından önerilen bootstrap prosedürleri kullanılması gösterilebilir. Bu prosedürlerde hata varyansları yapılan düzeltme formülleri ile daha yansız olarak kestirilmiştir.

Yapılan analizler sonucu varyans bileşenlerini kestirmede, açıklama yüzdeleri, mutlak ve bağıl hatalar ile standart hata değerleri göz önüne alındığında en uygun yöntemin $boot-m$ olduğu belirlenmiştir. Elde ettiğimiz bu sonuç Wiley (2001) ve Moore (2010) tarafından yapılan çalışmalar ile örtüşmekte iken, Brennan, Harris ve Hanson (1987) tarafından elde edilen sonuçlarla örtüşmemektedir. Brennan ve ark (1987)'na göre bootstrap için en uygun yöntem ANOVA yöntemi olarak belirtilmiştir. Ancak, bu farklılığın nedeni verilerin normal dağılımları ile ilgili varsayımlardan kaynaklanmaktadır. Brennan, Harris ve Hanson (1987)'un kullandığı veriler normal dağılım göstermekte iken, bu çalışmada ikili veriler kullanıldığından normal dağılımdan söz edilememektedir. Bu araştırmada da normal dağılıma sahip olmayan veriler kullanılmıştır. Veriler bootstrap yöntemi uygulandıktan sonra normal dağılım özellikleri göstermiştir.

G ve Φ katsayısı en büyük değerini bireylerin sabit tutulup maddelerin yeniden örneklemeye sokulduğu $boot-m$ prosedüründe almıştır. Otham (1995), tek yüzeyli G kuramı çalışmalarında, karar katsayıları hesaplamalarında maddelerin etkisinin diğer yüzeye göre daha fazla olduğunu belirtmiş ve $boot-m$ ile kestirilen Phi katsayısının en yüksek değerde çıkması gerektiğini belirtmiştir. Bu bakımdan araştırmada elde edilen değerler, önceki araştırmaları doğrular niteliktedir.

Kane (1996), evren puanı-hata oranlarının aynı çalışma üzerinde yapılacak değişkenliklerde daha anlamlı bir sonuç vereceğini belirtmiştir. Araştırmada değişkenlik olarak farklı varyans bileşenleri kestirme yöntemleri oluşturulmuştur ve bu yöntemlere ait değerler Tablo 4'te belirtilmiştir. Araştırmada en düşük evren puanı-hata değeri bağıl ve mutlak değerlendirmeler için maddelerin sabit tutulduğu, bireylerin yeniden örneklemeye sokulduğu $boot-b$ prosedüründe elde edilmiştir (S/N_{δ}=2,969, S/N_{Δ}=2,436). Bu sonuca göre $boot-b$ değerleri ile evren puanlarının varyansının hata varyansına olan büyüklüğü konusunda en iyi kestirimi sağladığı söylenebilir. Bir başka ifade ile elde edilen değer, örneklemden elde edilen gerçek puanlara karışan hata miktarının oranını belirler. Bu bakımdan evren puanları kullanılarak yapılan G çalışmalarında $boot-b$ ile kestirilen değerlerde genellenebilirlik evreni puanları daha iyi sağlıklı sonuçlar vereceklerdir. Fakat beklenenin aksine G ve Φ katsayıları değerlerine baktığımızda $boot-b$ yerine $boot-m$ prosedürü için en yüksek değerler elde edilmiştir. Bunun sebebi olarak da daha önce belirtildiği gibi örneklem sayısının azlığı gösterilebilir (Moore, 2010).

Araştırmada yeniden örnekleme sayısının, farklı büyüklüklerde alınması sonucu genellenebilirlik katsayıları ve indisler üzerine etkileri incelenebilir. Benzer bir araştırma birden çok yüzeye göre yapılabilir ve farklı varyans bileşeni kestirim yöntemleri hem çapraz hem de yuvalanmış desenler için ayrıntılı şekilde karşılaştırılabilir. İlerde yapılacak çalışmalarda gerçek veri seti veya gerçek veri setinden elde edilen parametreler kullanılarak varyans bileşenleri kestirim yöntemleri karşılaştırılabilir.

KAYNAKLAR

Brennan, R. L., Kane, M. T. (1977). An Index of Dependability for Mastery Tests. *Journal of Educational Measurement, 14*, 277-289.

Brennan, R. L., Harris, D. J., Hanson, B. A. (1987). The bootstrap and other procedures for examining the variability of estimated variance components in testing contexts. *ACT Research Report Series* 87-7. Iowa City, lA: American College Testing Program

Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer

Brennan, R. L. (2003*). Coefficients and Indices in Generalizability Theory*. (CASMA Research Report No.1). Iowa City: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.

Brennan R. L. (2007) Unbiased Estimates of Variance Components with Bootstrap Procedures. *Educational and Psychological Measurement, 67*, 784-803.

Cohen A. S., Kane M. T., Kim S. (2001). The Precision of Simulation Study Results. *Applied Psychological Measurement, 25*, 136-145.

Crocker, L., Algina, J. (1986). *Introduction to Classical and Modem Test Theory*. Fort Worth, FL: Harcourt Brace Jovanovich College Publishers

Cronbach, L. J., Rajarantnam, N., Gieser, G. C. (1963). Theory of Generalizability: A Liberalization of Reliability Theory. *British Journal of Statistical Psychology, 16*, 137-163.

Cronbach, L. J., Gieser, G. C., Nanda, H., Rajarantnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: Wiley

Efron, B., Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Hagvet K. A., Hoglend P. A. (2008). Assessing Precision of Change Scores in Psychodynamic Psychotherapy: A Generalizabiiity Theory Approach. *Measurement and Evaluation in Counseling and Development, 41*, 162-178.

Kane, M. T. (1996). The precision of measurements. *Applied Measurement in Education, 9,* 355–379.

Kane, M. (1999). *The Role of Generalizability in Validity.* Annual Meeting of the National Council on Measurement in Education. Montreal, Canada.

Leucht, R. M., & Smith, P. L. (1989). *The Effects of Bootstrapping Strategies on the Estimation of Variance Components.* Annual Meeting of the American Educational Research Association, San Francisco, California.

Meyer, J. P. (2010). *Reliability*. Oxford. Oxford University Press

Moore. J. L. (2010). *Estimating Standard Errors of Estimated Variance Components in Generalizability Theory Using Bootstrap Procedures*. Yayınlanmamış Doktora tezi, University of Iowa.

Othman, A. R. (1995). *Examining Task Sampling Variability in Science Performance Assessments*. Yayınlanmamış Doktora Tezi, University of California, Santa Barbara.

Rentz, J. O. (1987). Generalizability Theory: A Comprehensive Method for Assessing and Improving the Dependability of Marketing Measures. *Journal of Marketing Research*, *24*, 19-28.

Revelle, W. (2012). Package "psych": Procedures for Psychological, Psychometric, and Personality Research. Version: 1.2.4. <http://cran.r-project.org/web/packages/psych/psych.pdf>

Searle, S. R. (1987*), Linear Models for Unbalanced Data*, New York: John Wiley & Sons Publications.

Shavelson, J. R. ve Webb N. M. (1991). *Generalizability Theory: A Primer.* Newbury Park. CA: Sage Publications.

Shavelson, R. J. ve Webb, N. M. (2004). G*eneralizability Theory. Encyclopedia of Social Measurement.* New York: Academic Press.

Wiley, E. W. (2001). *Bootstrap strategies for variance component estimation: Theoretical and empirical results.* Yayınlanmamış Doktora tezi, Stanford University.

EXTENDED ABSTRACT

Introduction

The initial definiton of the Generalizability theory was provided by Cronbach and his friends (1963) which is response to the limitations of reliability estimations based on error term which is simply the difference between true score and observed score in classical test theory. The main purpose of generalizability theory is to decompose the total observed score into different variance components, so that it become easy to interpret, define and generalize the individuals observed scores to universe of generalization correctly. Generalizability theory provides an extensive conceptual framework and set of statistical procedures for quantifying and explaining the consistencies and inconsistencies in observed scores for objects of measurement.

In generalizability theory, variance components assume central importance. Numerous procedures might be used to estimate variance components, but by far the most frequently employed procedure in generalizability theory is the so-called "ANOVA" procedure that involves equating mean scores to their expected values, and then solving for the estimated variance components. Searle (1987) made significant contrubitions to the field that studying various variance estimation procedures called minimum norm quadratic unbiased estimation (MINQUE), maximum likelihood (ML), restricted maximum likelihood (REML). Later on researchers started to study on error based estimated variance components called bootstrap (Brennan, Harris ve Hanson, 1987; Luecht ve Smith, 1989; Shavelson ve Webb, 2004). In non-normal distrubitions, researchers needed a method which data don't require normality assumptation (Brennan, 2001). In that case variance components would be estimated for each of a given number of bootstrap samples, and the standard deviation of each variance component estimate would be calculated across the bootstrap samples as an estimate of the standard error of the variance component. An advantage of the bootstrap is that it does not require assumptions about the distribution of the statistic.

The aim of this research is to compare various variance component estimations procedures with using signal noise ratio and error tolerance ratio which is offered with generalizabiity and Phi coefficients in non-normal distrubitions (Brennan, 2001; Kane, 1999). This research compares variance components estimations with using ANOVA and bootstrap procedures in non-normal disturbitions in one facet corssed design G studies.

Method

Data's were gathered by using two seperate procedures (a) data simulation and (b) sampling simulation. In data simulation part, it's been simulated a non-normal dichotomous data set which fits to unidimensional person-item matrix 60x5 which fits to $b \times m$ design. All the simulations replicated 25 times. In sampling simulation sections datas, gathered from data simulation sections has been bootstrapped 1000 times according to the each facet.

Results and Discussion

Standart errors, variance components, relative and absolute error are estimated according to the each facets with using ANOVA and bootstrap procedures. It's been found that lowest absolute and relative standart errors estimated in $boot-m$ procedure (SH(δ)=0,109), and highest absolute and relative standart errors estimated in ANOVA procedure (SH(δ)=0,113). According to the variance estimation analysis, $boot-m$ has the best estimations and referred as preferable method regarding relative and absolute errors and standart errors. It's been found that highest values of G and Φ coeffieicnts estimated in $boot-m$ procedure in which individuals are stable and items are on bootstrapping. The results also show that in non-normal dichotomously scored datas best signal-noise ratio has estimated in $boot-b$ procedure, and best error-tolerance ratio has been estimated in $boot-m$ procedure. Thus, $boot-m$ procedures gives more valid estimations and $boot-b$ procedure gives more reliable and precise estimations of universe scores in G studies rather than other procedures.