Yinelenen ölçümler (/*Repeated Measures*) Analizinde Küresellik (/*Sphericity*) ve Bileşik Simetri (/*Compound Symmetry*)

Prof.Dr. Ömer Cevdet Bilgin
Atatürk Üniversitesi

Aynı deneklerin (/*subjects*) bir faktörün çok sayıdaki düzeyi altında test edildiği yinelenen ölçümler dizaynlarının pek çok pratik ve istatistiksel faydası vardır. Bunlardan birisi, gruplar-arası bireysel farklılıkların neden olduğu hata varyansını azaltmasıdır. Mamafih, hatadaki bu azalmanın şu bedeli vardır ki yinelenen ölçümler dizaynları potansiyel olarak muameleler arasında kovaryasyona sahiptir. Çünkü, aynı birey her muamelede kullanılır ve böylece muameleler arasında bireylerin davranışları muhtemel bir tutarlılık gösterir. Gruplar-arası ANOVA da, test edilen grupların bağımsız olduğu varsayımı yürürlükte olmak zorundadır ki yapılan test isabetli (/*accurate*) olsun. Scariano ve Davenport (1987) bu varsayımın zorlanmasının bazı vargılarını (/*consequences*) yazmışlardır. Şöyle ki, yinelenen ölçümler dizaynında muameleler arasındaki ilişki test istatistiğinin isabetliliğine dair problemler yaratır. Bu metnin gayesi yinelenen ölçümler verisinin ANOVA ile analizinde ortaya çıkan bu mevzuları olanaklıca basitçe açıklamaktır: Daha spesifik olarak, küresellik (/*sphericity*) nedir ve niçin önemlidir?

Küresellik Nedir?

Derslerden şunu anımsarız ki farklı deneklere ait veri analiz edilirken, muameleler arasında varyansların homojenliği çok önemlidir. Fakat, çoğu zaman, yinelenen ölçümler dizaynlarında bu problem sanki ortadan kalkmış gibi davranılır. Böyle değil elbette. ANOVA daki varyansların homojenliği varsayımının yerini yinelenen ölçümlerin analizinde küresellik varsayımı alır.

Küresellik ε ile temsil edilir ve bazen *sirkülerlik* olarak ta adlandırılır. *Bileşik simetrinin* daha genel bir formudur. Bir Σ popülasyon varyans-kovaryans matrisine sahip olduğumuzu farzedelim.

$$\Sigma = \begin{bmatrix} s_{11}^2 & \alpha_{12} & \alpha_{13} & \dots & \alpha_{1n} \\ \alpha_{21} & s_{22}^2 & \alpha_{23} & \dots & \alpha_{2n} \\ \alpha_{31} & \alpha_{32} & s_{33}^2 & \dots & \alpha_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} & \dots & s_{nn}^2 \end{bmatrix}$$

Bu matris şu iki şeyi ihtiva ediyor: (1) Diyagonal elemanlar ($s_{11}^2, s_{12}^2, \dots, s_{nn}^2$) ki n adet muamelenin her biri içerisindeki varyansları temsil etmekte. (2) Diyagonal-dışı elemanlar ($\alpha_{12}, \alpha_{13}, \dots, \alpha_{nn-1}$) ki her iki muamele arasındaki kovaryansları temsil etmekte (bunu ölçümlerin yinelendiği muameleler arasında standartlaştırılmamış korelasyon olarak düşünebilirsin).

Muameleler arasında varyansların homojenliği varsayımının yürürlüğü şuna bağlıdır:

$$s_{11}^2 \approx s_{22}^2 \approx s_{33}^2 \approx \ldots \approx s_{nn}^2$$

Yani, matrisin diyagonal elemanlarının yaklaşıkça eşit olmaları gerekir. Bu bir gruplar-arası dizaynda beklediğimiz durumla mukayese edilebilir. Mamafih, yinelenen ölçümler dizaynlarında, ilave bir komplikasyon vardır. Şöyle ki muameleler biribirleriyle kovaryanslıdır. Böyle olunca, veriyi analiz ederken bu kovaryansların etkisini hesaba katmalı ve spesifik olarak bu kovaryansların tümünün yaklaşıkça eşit olduğunu (yani, muamelelerin tümünün biribirleriyle aynı derecede ilişkili olduğunu ve böylece bir muamelede yer aldıktan sonra bir başka muamelede yer almanın etkisinin de eşit olduğunu) varsaymalıyız. Matrisin diyagonalinde sabit varyansların (yani, varyansların homojenliği) ve diyagonal-dışında sabit kovaryansların (yani, muameleler arasındaki kovaryansların eşitliği) bir biçimi (/*pattern*) olduğu zaman bileşik simetri yürürlüktedir.

$$\alpha_{12} \approx \alpha_{13} \approx \alpha_{23} \approx \ldots \approx \alpha_{1n} \approx \alpha_{2n} \approx \alpha_{3n} \approx \ldots$$

Küresellik bileşik simetrinin daha az kısıtlayıcı bir formudur (gerçekte, yinelenen ölçümler ANOVA ile ilgili pek çok araştırmada bileşik simetri ve küresellik hakkında kafa karışıklığı mevcuttur). Küresellik faktör düzeyleri arasındaki farkların varyanslarının eşitliğine matuftur. Bileşik simetri muameleler arasındaki kovaryasyonla alakadar iken, küresellik muameleler arasındaki farkların varyansına ilişkindir. Buna göre, eğer faktör düzeylerinin her bir çifti için skor çiftleri arasındaki farkları hesaplarsak, o zaman bu farkların varyanslarının eşit olması lazımdır. Farzedelim ki bir yinelenen ölçümler faktörünün 4 düzeyi (A, B, C, D) olsun. Küreselliğin yürürlükte olması için, şu eşitlik sağlanmalıdır:

$$s_{A-B}^2 \approx s_{A-C}^2 \approx s_{A-D}^2 \approx s_{B-C}^2 \approx s_{B-D}^2 \approx s_{C-D}^2$$

Yukarıdaki eşitlik sağlanmadığı (yani, muamele çiftleri arasındaki farkların varyanslarının eşit olmaması) zaman, küresellik varsayımı zorlanmış olur.

Küresellik Nasıl Ölçülür?

Küresellik varsayımının karşılanıp karşılanmadığını görmenin en basit yolu faktör düzeylerinin tüm ikili kombinasyonlarına ait skor çiftleri arasındaki farkları hesaplamaktır. Bunu yaptıktan sonra, bu farkların varyanslarını basitçe hesaplayabiliriz. Mesela müteakip tablonun göseterdiği hipotetik veri bir faktörün 3 düzeyi ile yapılan bir deneye aittir (basitleştirmek adına, her muamele için yalnızca 5 skor verilmiştir).

A Muamelesi	B Muamelesi	C Muamelesi	A-B	A-C	B-C
10	12	8	-2	2	5
15	15	12	0	3	3
25	30	20	-5	5	10
35	30	28	5	7	2
30	27	20	3	10	7
		Varyans	**15.7**	**10.3**	**10.3**

Faktör düzeylerinin çiftleri arasındaki farklar her bir denek için hesaplanabilir. Farkların her bir kümesinin varyansı hesaplanabilir. Yukarıda dediğimize göre, eğer bu varyanslar kabaca eşitse, küresellik varsıyımı yürürlükte olacaktır. Bu veri için, küresellik şu halde yürürlüktedir:

$$s^2_{A-B} \approx s^2_{A-C} \approx s^2_{B-C}$$

Ki burada:

$$s^2_{A-B} = 15.7$$
$$s^2_{A-C} = 10.3$$
$$s^2_{B-C} = 10.3$$

Buna göre,

$$s^2_{A-B} \neq s^2_{A-C} = s^2_{B-C}$$

Görüdüğümüz üzere, en azından küresellikten bir sapma var. Çünkü, A muamelesi ve B muamelesi arasındaki farkların varyansı A ve C arasındaki farkların ve B ve C arasındaki farkların varyanslarından daha büyük. Mamafih, bu verinin *lokal sirkülerliğe* (veya lokal küreselliğe) sahip olduğunu söyleyebiliriz, çünkü varyanslardan ikisi özdeştir. Bunun manası şudur ki bu farkları ihtiva eden çoklu karşılaştırmalar (/*multiple comparisons*) için, küresellik varsayımı karşılanmaktadır (lokal sirkülerlik için Rouanet ve Lepine, 1970 e bakılabilir). Yukarıdaki tablodaki verinin küresellikten sapması fazla şiddetli görünmüyor (varyansların tümü *kabaca* eşit). Bu şu meseleyi ortaya koyar ki: Eylemimizi garantilemek için, küresellik varsayımını zorlamaların yeterince şiddetli olup olmadığını nasıl değerlendireceğiz?

Küresellikten Ayrılışların Şiddetini Değerlendirme

Kompüter paketlerindeki ilerlemeler sayesinde küresellikten ayrılışları değerlendirmek için detaylarla uğraşmaya ihtiyacımız yok artık. SPSS Mauclhley testi olarak bilinen bir test

istatistiği üretir öyle ki muameleler arasındaki farkların varyanslarının eşit olduğu hipotezini test eder. Buna göre, eğer Mauchley test istatistiği anlamlı ise (yani 0.05 ten daha küçük bir olasılık değerine sahipse), şu neticeye varmalıyız ki farkların varyansları arasında anlamlı farklar vardır. Buna binaen, küresellik koşulu karşılanmamıştır. Mamafih, eğer Mauchley test istatistiği anlamsız ise (yani 0.05 ve daha büyük bir olasılık değerine sahipse), o zaman şu neticeye varmak makuldür ki farkların varyansları anlımlıca farklı değildir (yani kabaca eşittir). Öyleyse, kısaca, eğer Mauchley testi anlamlıysa, o zaman kompüterin ürettiği *F*-oranları hakkında endişelenmeliyiz.

Aşağıdaki figür SPSS v7.0 çıktısında üç muamele (A, B ve C) için uydurulmuş bir verinin Mauchley testi sonucunu gösteriyor.

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
FACTOR1	.011	13.485	2	.001	.503	.506	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: FACTOR1

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the layers (by default) of the Tests of Within Subjects Effects table.

Testin sonucu anlamlı ve şunu belirtiyor ki farklar arasındaki varyanslar anlamlıca farklıdır. Tablo aynı zamanda serbestlik derecelerini (/*df: degrees of freedom*, basitçe $n-1$ dir ve burada n karşılaştırılan varyansların sayısıdır) ve küreselliğe ait üç kestirimi (küreselliği düzeltmeyi metnin devamında ele alacağız) de gösteriyor. Serbestlik derecesini bu metinin devamında *sd* ile temsil edeceğiz.

Küresellik Varsayımını Zorlamanın Etkisi Nedir?

Küresellik varsayımı yürürlükte olmadığı zaman, *F*-testinin geçerliliği Rouanet ve Lepine (1970) tarafından detaylıca tartışılmıştır. İleri sürdüler ki muamele karşılaştırmalarını değerlendirmek için kullanılabilecek iki farklı *F*-oranı vrdır. Bu iki tip *F*-oranı srasıyla F' ve F'' ile simgelenir. F' sorgulanan karşılaştırmanın kareler aritmetik-ortalaması ve deneklerin o karşılaştırma ile interaksiyonlarından türetilen bir *F*-oranına matuftur. F'', spesifik hata kareler aritmetik-ortalamasından değil, yinelenen ölçümler karşılaştırmalarının tümüne ait total hata kareler aritmetik-ortalamasından türetilir. Rouanet ve Lepine (1970) yine ileri sürdüler ki F'' ne göre F' daha az güçlüdür ve bu yüzden bu test istatistiği hakiki etkileri ıskalayabilir. İlaveten, gösterdiler ki F' nün geçerli olabilmesi için Σ varyans-kovaryans matrisi lokal sirkülerliğe tabi olmalıdır (yani, sorgulanan spesifik karşılaştırma için küresellik yürürlükte olmalıdır). Mendoza, Toothaker ve Crain (1976) bunu destekledilerler. F'' yalnızca *genel sirkülerliği* gerektirir (yani, verinin tümü sirküler olmalıdır) fakat sirkülerliğin ve bileşik simetrinin resiprokal-olmayan mahiyeti icabı, F'' bileşik simetriyi gerektirmez, oysa F' gerektirir. Böylece, genel olarak kullanılan istatistik F' olduğuna göre, küreselliğin zorlanmasının etkisi bir güç kaybı (F'' nün yerine kullanıldığı zaman) ve geçersiz olabilecek bir test istatistiğidir (*F*-oranı).

Küresellik Zorlamaları İçin Düzeltme

Eğer veri küresellik varsayımını zorluyorsa, geçerli bir *F*-oranı üretmek için uygulanabilecek bir kaç düzeltme vardır. SPSS üç düzeltme hesaplar ki bunlar Greenhouse ve Geisser (1958) ve Huynh ve Feldt (1976) tarafından tavsiye edilen küresellik kestirimlerine dayanır. Bu kestirimlerin her ikisi de gözlenen *F* değerini elde etmek için kullanılan serbestlik derecelerine uygulanan birer düzeltme faktörü temin eder. Bu kestirimlerin nasıl hesaplandıkları bu metnin kapsamı dışındadır. Kendi maksadımıza matuf olarak, bu kestirimlerin biribirinden azca farklı olduğunu bilmemiz kafidir. Greenhouse-Geisser kestirimi (genellikle $\hat{\varepsilon}$ ile temsil edilir) $1/k-1$ (ki burada k ölçümlerin yinelendiği muamelelerin sayısıdır) ve 1 arasında değişir. $\hat{\varepsilon}$ nın değeri 1 e ne kadar yakınsa, farkların varyansları o kadar homojendir ve dolayısıyla veri küreselliğe o kadar yakındır. Yukarıdaki figür üç muameleli bir durumu gösteriyor ve bu yüzden $\hat{\varepsilon}$ nın alt limiti 0.5 . Görüyoruz ki $\hat{\varepsilon}$ nın hesaplanan değeri 0.503. Bu da 0.5 e çok yakın ve küresellikten esaslı bir sapmayı temsil ediyor. Hyunh ve Feldt (1976) rapor etmişlerdir ki $\hat{\varepsilon} > 0.75$ olduğu zaman çok sayıda yanlış sıfır hipotezi ret edilemeyecektir (yani, test fazla tutucudur) ve Collier, Baker, Mandeville ve Hayes (1967) göstermişlerdir ki bu aynı zamanda $\hat{\varepsilon}$ nın 0.9 kadar büyük olduğu zamanlarda da doğrudur. Huynh ve Feldt, bundan ötürü, $\hat{\varepsilon}$ yı daha az tutucu yapmak için bir düzeltme önerdiler (genellikle $\tilde{\varepsilon}$ ile simgelenir). Mamafih, Maxwell ve Delanay (1990) rapor ettiler ki $\tilde{\varepsilon}$ küreselliği fiilen aşırı-kestiriyor. Stevens (1992) bundan ötürü bu iki değerin ortalamasını almayı ve *sd* ni bu ortalama değere göre düzeltmeyi tavsiye etti. Girden (1992) tavsiye etti ki $\hat{\varepsilon} > 0.75$ olduğu zaman *sd* ni düzeltmek için $\tilde{\varepsilon}$ kullanılmalı; eğer $\hat{\varepsilon} < 0.75$ ise veya küresellik hakkında hiç bir şey bilinmiyor ise, o zaman *sd* ni düzeltmek için tutucu $\hat{\varepsilon}$ kullanılmalı.

Aşağıdaki füğür SPSS v.7.0 çıktısında küreselliği zorlayan bir veri kümesine ait tipik ANOVA tablosunu ve ε-düzeltilmiş *F*-değerlerini gösteriyor (veri kümesi yukarıdaki tabloyu üretmek için kullanılan ile aynı).

Tests of Within-Subjects Effects

Measure		Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
MEASURE_1	Sphericity Assumed	FACTOR1	2895.600	2	1447.800	5.245	.035	10.489	.662
		Error(FACTOR1)	2208.400	8	276.050				
	Greenhouse-Geisser	FACTOR1	2895.600	1.006	2879.437	5.245	.083	5.274	.418
		Error(FACTOR1)	2208.400	4.022	549.018				
	Huynh-Feldt	FACTOR1	2895.600	1.011	2863.394	5.245	.083	5.304	.420
		Error(FACTOR1)	2208.400	4.045	545.959				
	Lower-bound	FACTOR1	2895.600	1.000	2895.600	5.245	.084	5.245	.417
		Error(FACTOR1)	2208.400	4.000	552.100				

a. Computed using alpha = .05

Tablo bize *F*-değerini ve küresellik varsayımı altında ilgili serbestlik derecesini veriyor. Buna göre, anlamlı bir *F*-istatistiği muamelelerin aritmetik-ortalamaları arasında bir fark olduğunu belirtiyor. Aşağısında, düzeltilmiş değerler veriliyor (küreselliğin her üç kestirimi için). Her üç durumda da *F*-oranlarının aynı kaldığını görüyoruz; değişen serbestlik dereceleri (ve dolayısıyla *F* nin kritik değeri). Serbestlik derecesi küresellik kestirimi ile düzeltiliyor. Bunun nasıl yapıldığını şöyle gösterebiliriz:

Kullanılan küresellik kestirimi	Kestirimin değeri	Terim	*sd*	Düzeltme	Yeni *sd*
Yok		Etki	2		
		Hata	8		
	0.503	Etki	2	0.503×2	1.006
		Hata	8	0.503×8	4.024
	0.506	Etki	2	0.506×2	1.012
		Hata	8	0.506×8	4.048

Çokdeğişkeli ve Tekdeğişkeli Testler

Elimizeki veri küreselliği zorladığı zaman, son bir opsiyonumuz. Çokdeğişkeli test istatistiklerini (MANOVA) kullanmaktır, çünkü küresellik varsayımı altında değillerdir (O'brien ve Kaiser, 1958 e bakılabilir). Tekdeğişkeli ve çokdeğişkeli yaklaşımlar arasında bir güç takası söz konusu olmasına rağmen, bazı yazarlar şunu ileri sürerler ki bu tekniklerin ustaca kullanılması suretiyle, bunun üstesinden gelinebilir (O'brien ve Kaisser, 1985). MANOVA küresellik varsayımını göz ardı eder (ve mütenasip *F*-oranları ve düzeltmeleri hakkındaki mütekabil mülahazaların tümünü). Onun yerine, 1 *sd* li karşılaştırmalar için spesifik bir hata terimi kullanmak suretiyle, her bir karşılaştırmayı yalnızca kendi spesifik hata terimi ile ilişkilendirir (ANOVA da kullanılan karma hata terimi yerine). Davidson (1972) düzeltilmiş tekdeğişkeli tekniklerin gücünü Hotelling T^2 ninki ile karşılaştırdı (bir MANOVA test istatistiği) ve şunu buldu ki düşük korelasyonlu muamelerle beraber mevcut oldukları zaman, yüksek korelasyonlu muameleler arasındaki küçük mazbut değişiklikleri (/*changes*) tespit etmekte, tekdeğişkeli teknikler nispeten güçsüzdür. Mendoza, Toothaker ve Nicewander (1974) bileşik simetri ve normallik zorlamaları altında tekdeğişkeli ve çokdeğişkeli teknikleri karşılaştıran bir Monte Carlo öğrenme-çalışması yürüttüler ve buldular ki bileşik simetri varsayımı daha yüksek derecede zorlandıkça, çokdeğişkeli testlerin ampirik gücü artarken, tekdeğişkeli testlerin gücü azalmaktadır. Maxwell ve Delanay (1990) yazmışlardır ki n küçülürken tekdeğişkeli test çokdeğişkeli testten nispeten daha güçlüdür ve şunu önermişlerdir ki n eğer $a + 10$ dan daha küçükse (a yinelenen ölçümlere ait muamelelerin sayısıdır), çokdeğişkeli yaklaşım muhtemelen kullanılmamalıdır. Genel bir kural olarak, öyle görünüyor ki elimizdeki veri küresellik varsayımını şiddetlice zorluyorsa ($\varepsilon < 0.7$) ve örnek hacmimiz ($a + 10$) dan daha geniş ise, o zaman çokdeğişkeli prosedürler daha güçlüdür. Öte taraftan, örnek hacmimiz dar ise veya küresellik yürürlükteyse ($\varepsilon > 0.7$) ise, tekdeğişkeli yaklaşım tercih edilir (Stevens, 1992). Şunu da ilave etmekte fayda var ki MANOVA nın gücü, yanıt değişkenleri arasındaki korelasyonların bir fonksiyonu olarak, artar veya azalır (Cole ve ark., 1994). Demek ki aynı zamanda muameleler arasındaki ilişki de mülahaza edilmelidir.

Çoklu Karşılaştırmalar

Şimdiye kadar, küreselliğin omnibüs ANOVA üzerindeki etkisini tartıştık. Son olarak çoklu karşılaştırma prosedürleri üzerindeki etkilerini de kısaca tartışacağız. Boik (1981) küreselliğin

yinelenen ölçümler dizaynlarında *a priori* testler üzerindeki etkilerine ait dikakte değer bir muhasebe yapmış ve şu neticeye varmıştır ki küresellikten çok küçük ayrılışlar bile *F*-testinde büyük yanlılıklara yol açmaktadır ve tavsiye etmiştir ki yinelenen ölçümlere ait kontrastlar için bu testler kullanılmasın. Deney hatası terimleri küçük olduğu zaman, nispeten büyük miktardaki etkileri tespit etme gücü 0.05 e kadar azalabilmektedir (küresellik 0.8 olduğu zaman). *A priori* karşılaştırmalar için, yazar bu durumun iyileştirilemeyeceğini öne sürmüş ve neticede bir çokdeğişkeli analogunu tavsiye etmiştir. Mitzel ve Games (1981) buldular ki küresellik yürürlükte olmadığı zaman ($\varepsilon < 1$), eşli karşılaştırmalarda (/*pairwise comparisons*) geleneksel olarak istihdam edilen karma hata terimi (/*pooled error term*) iki aritmetik-ortalama arasında anlamsız farklar ile veya tespit edilemeyen farklar (tutucu bir I. Tip hata oranı) ile sonuçlanmaktadır. Bu sebeple, her bir karşılaştırma için ayrı bir hata teriminin kullanılmasını tavsiye ettiler. Maxwell (1980) yinelenen ölçümler muameleleri için 5 ayrı *a priori* testin gücünü ve α-düzeylerini sistematik olarak test etti. Değerlendirilen testler şunlardı: karma bir hata terimi kullanan Tukey Bütüncül Anlamlı Fark (WSD) testi, $(n-1)$ sd'li ve $(n-1)(k-1)$ sd'li hata terimlerini kullanan iki farklı Tukey prosedürü (SEP1 ve SEP2), Bonferroni prosedürü (BON) ve Roy-Bose Simültane Güven Aralığı (SCI) adlı bir çokdeğişkeli yaklaşım. Maxell bu *a priori* prosedürleri değişen örnek hacimleri, yinelenen faktörün düzeyleri ve küresellikten ayrılışlar ile test etti. Şunu buldu ki çokdeğişkeli yaklaşım "pratikte kullanmak için çok fazla tutucudur" ve bu n in (deneklerin sayısı) k ya (muamelelerin sayısı) nispetle küçük olduğu durumlarda en aşırıdır. Varyans-kovaryans matrisi küresellikten ayrılırken, Tukey testi α-oranını artırdı ve hatta ayrı bir hata terimi (SEP1) kullanıldığı zaman, k artarken, α hafifçe arttı. Yine SEP2 kabul-edilebilmezce yüksek α düzeylerine yol açtı. Bonferroni metodu, mamfih, aşırıca sağlamdı (/*robust*) (*hafifçe* tutucu olmakla beraber) ve hangi manipülasyon yapılırsa yapılsın α düzeyleri kontrol altında tutuldu. Bundan ötürü, I. Tip hata oranları bakımından, Bonferroni metodu en iyisiydi. Küçük bir örnek ($n = 8$) için testin gücü (II. Tip hata oranı) bakımından, küreselsizlik (/*nonsphericity*) koşulları altında, WSD en güçlüsüydü. Bu avantaj, $n = 15$ iken şiddetlice azaldı. Keselman ve Keselman (1988) Maxwell'in çalışmasını genişlettiler ve aynı zamanda dengesiz (/*unbalanced*) dizaynları da soruşturdular. Bu kişiler, yine, Tukey WSD, modifiye edilmiş bir WSD (karılmamış hata varyansıyla), Bonferroni *t*-istatistikleri ve bir çokdeğişkeli yaklaşım kullandılar ve Maxwell'inkilerle aynı faktörlere (eşitsiz örnekleri de ilave ederek) baktılar. Şunu buldular ki ağırlıklandırılmamış aritmetik—ortalamalar (dengesiz dizaynlar ile) kullanıldığı zaman, dört testin hiç biri I. Tip hata oranını kontrol altında tutabilmedi. Ağırlıklandırılmış aritmetik-ortalamalar kullanıldığı zaman, yalnızca çokdeğişkeli testler α oranlarını sınırlandırabilirken, öte tatarftan iki Tukey metoduna göre, Bonferroni *t* istatistikleri dikkate değerce daha iyiydi. Güç bakımından, bu yazarlar şu neticeye vardılar ki "yinelenen ölçümler muamelelerinin sayısı artarken, SCI ye göre, BON esaslıca daha güçlüdür".

Anılan öğrenme-çalışmalarına göre, Bonferroni metodu tekdeğişkeli tekniklerin genelce en sağlamı olarak görünmektedir, bilhassa güç ve I. Tip hata oranının kontrol altına alınması bakımından.

Netice

İstisna olmaktan daha ziyade çoğu zaman bir kural olarak, yinelenen ölçümler dizaynlarında küresellik zorlanır. Bu sebeple, tüm yinelenen ölçümler dizaynları küresellik zorlaması testlerine tabi tutulmalıdır. Eğer küresellik zorlanıyorsa, o zaman araştırmacı bir çokdeğişkeli

test mi yoksa bir tekdeğişkeli test mi tercih edeceğine (bir yanda testin geçerliliği ve bir yanda testin gücü arasında bir takas mülahazasıyla) karar vermelidir. Eğer tekdeğişkeli metotlar tercih edilirse, o zaman omnibüs ANOVA, küresellikten ayrılışın düzeyine bağlı olarak, mütenasipçe düzeltilmelidir.

Son olarak, eğer eşli karşılaştırmalar gerekliyse, I. Tip hata oranını kontrol altına almak için, Bonferroni metodu muhtemelen kullanılmalıdır. Ancak, grup hacimlerinin eşit olduğundan emin olmak gerekir, aksi halde, Bonferroni teknikleri dahi α düzeylerinin artmasına maruz kalır.

Referans

Boik, R. J. (1981). A Priori tests in repeated measures designs: effects of nonsphericity, *Psychometrika*, **46** (3), 241-255.

Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1994). How he power of MANOVA can both increase and decrease as a function of the intercorrelations among the dependent variables, *Psychological Bulletin*, **115** (3), 465-474.

Field, A. (2009). Discovering Statistics Using SPSS. 3rd Ed. SAGE.

Girden, E. R. (1992). *ANOVA: Repeated Measures* (Sage university paper series on qualitative applications in the social sciences, 84), Newbury Park, CA: Sage.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, **24**, 95–112.

Huynh, H., and Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomised block and split-plot designs. *Journal of Educational Statistics*, **1** (1), 69-82.

Keselman, H. J. & Keselman, J. C. (1988). Repeated measures multiple comparison procedures: Effects of violating multisample sphericity in unbalanced designs. *Journal of educational Statistics*, **13** (3), 215-226.

Maxwell, S. E. (1980). Pairwise multiple comparisons in repeated measures designs. *Journal of Educational Statistics*, **5** (3), 269-287.

Maxwell, S. E. & Delaney (1990). *Designing experiments and analyzing data*. Belmont, CA: Wadsworth.

Mendoza, J. L., Toothaker, L. E. & Nicewander, W. A. (1974). A Monte Carlo comparison of the univariate and multivariate methods for the groups by trials repeated measures design. *Multivariate Behavioural Research*, **9,** 165-177.

Mendoza, J. L., Toothaker, L. E. & Crain, B. R. (1976). Necessary and sufficient conditions for F Ratios in the L x J x K Factorial design with two repeated factors. *Journal of the American Statistical Association*, **71,** 992-993.

Mitzel, H. C., & Games, P. A. (1981). Circularity and multiple comparisons in repeated measures designs, *British Journal of Mathematical and Statistical Psychology*, **34**, 253-259.

O' Brien, M. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer, *Psychological Bulletin*, **97** (2), 316-333.

BPS-MSC Newsletter 6 (1) Page 24

Rouanet, H. & Lépine, D. (1970). Comparison between treatments in a repeated-easurement design: Anova and multivariate methods. *The British Journal of mathematical and Statistical Psychology*, **23,** 147-163.

Scariano, S. M. & Davenport, J. M. (1987). The effects of violations of independence in the oneway ANOVA. *The American Statistician*, **41** (2), 123–129.

Stevens, J. (1992). *Applied multivariate statistics for the social sciences* (2nd edition). Hillsdale, NJ: LEA.