DERLEME / REVIEW

Düzce Üniversitesi
Sağlık Bilimleri Enstitüsü Dergisi
2014;4(1): 28-33
ISSN: 2146-443X

sbedergi@duzce.edu.tr

Birinci Tür Hata'nın Kontrolü ve Adımsal (Stepwise) Çoklu Karşılaştırma Testleri

Nurhan DOĞAN[1], İsmet DOĞAN[1]

ÖZET
Varyans Analizi, ikiden çok ortalamanın karşılaştırılmasında, ortalamalar arasında fark olup olmadığını belirlemekte, ancak bu farklılığın nereden kaynaklandığını belirleyememektedir. Çoklu karşılaştırma yöntemleri ortalamalar arasındaki farklılıklar hakkında çok daha detaylı bilgi vermektedir. Bu amaçla çok sayıda çoklu karşılaştırma testi önerilmiştir. Ancak hangi durumda hangi testin kullanılacağına karar vermek kolay değildir. Bu çalışmada amaç, son yıllarda daha yoğun olarak kullanılmaya başlanan ancak araştırmacılar tarafından daha az bilinen yedi farklı adımsal çoklu karşılaştırma testlerini tanıtmaktır.
Anahtar Kelimeler: Varyans analizi; birinci tür hata; adımsal çoklu karşılaştırma işlemleri.

The Control of Type I Error Rate and Stepwise Multiple Comparison Procedures: A Review

ABSTRACT
Analysis of variance determines whether the difference between averages when more than two averages compare, but it cannot determine origin of this difference. Detailed information about the difference between means is derived by using multiple comparison tests. Many different tests have been suggested for this purpose. The present article reviews of stepwise multiple comparison procedures. The review was motivated by examples of multiple comparison practices. Making a decision about which test to use is not easy task. The aim of this study is to introduce seven different stepwise multiple comparison tests which are less known by the researchers.
Keywords: Analysis of variance; type I error rate; stepwise multiple comparison procedure.

GİRİŞ VE AMAÇ

Varyans analizi, ikiden çok ortalamanın karşılaştırılmasında, ortalamalar arasında fark olup olmadığını belirlemekte, ancak farklılığın kaynağını belirleyememektedir. Çoklu karşılaştırma yöntemleri ise üç ya da daha fazla denemenin hangisinin daha iyi hangisinin daha kötü olduğuna karar verebilmek, ortalama etkilerini karşılaştırmak, yanlış bir karar verme olasılığının kontrol edilmesinde hangi denemenin ne kadar daha iyi ya da ne kadar daha kötü olduğuna karar vermektir.

Rao ve Swarupchand (1)'a göre çoklu karşılaştırma terimi bir grup içindeki varyanslar, oranlar ya da ortalamalar arasındaki farklılığın istatistiksel anlamlılığı için yapılan testleri ifade etmektedir. Çoklu karşılaştırma yöntemleri, çokluluk etkisinden kaynaklanan hatalı sonuç çıkarmaların düzenli kontrolünü göz önünde bulunduran istatistiksel işlemlerdir. Çoklu karşılaştırma yöntemleri uygulamalardaki önemliliğinden dolayı temel bir problemdir ve farklı yollarla kullanıcılara yol göstermektedirler. Araştırmacıların amacına göre literatürde görülen dört tip çoklu karşılaştırma yöntemi söz konusudur. Bunlar;

- Tüm ikili çoklu karşılaştırmalar (**a**ll-pairwise **m**ultiple **c**omparison (MCA), $i \neq j$ olmak üzere tüm μ_i-μ_j farklarının karşılaştırılması dikkate alınmaktadır),
- En iyi ile çoklu karşılaştırmalar (**m**ultiple **c**omparison with the **b**est (MCB), $i \neq j$ ve i=1,2,…,k olmak üzere tüm μ_i-maxμ_j farklarının karşılaştırılması dikkate alınmaktadır),

[1] Afyon Kocatepe Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı - AFYONKARAHİSAR

Correspondence: Dr. Nurhan DOĞAN, e-posta: ndogan@aku.edu.tr

- Bir kontrol ile çoklu karşılaştırmalar (**m**ultiple **c**omparison with a **c**ontrol (MCC), i=1,2,…,k-1 olmak üzere tüm μ_i-μ_k farklarının karşılaştırılması dikkate alınmaktadır),
- Ortalama ile çoklu karşılaştırmalar (**m**ultiple **c**omparison with the **m**ean (MCM), i=1,2,…,k olmak üzere tüm μ_i -$\bar{\mu}$ farklarının karşılaştırılması dikkate alınmaktadır).

Çoklu karşılaştırma konusunun istatistiğin en çok karışıklık yaratan konulardan biri olduğu ve bu konu ile ilgili literatürde oldukça farklı önerilerin bulunduğu ifade edilmektedir (2). Halen bir grup ortalamanın çoklu karşılaştırma işlemleri gerçekleştirilirken hangi yöntemin hangi durumda kullanılması gerektiği konusunda görüş birliği bulunmamaktadır. Araştırmacılar, biraz farklı gerekçelerle de olsa kendi tercihlerini yaparak benzer problemlerin çözülmesinde farklı çoklu karşılaştırma işlemleri kullanmaktadırlar (3).

Araştırmacılar, literatürde yer alan klasik çoklu karşılaştırma testlerini kullanırlarken gerçekte çoklu karşılaştırma testlerinin kullanılmasının bir sonucu olarak ortaya çıkan birinci tür hatanın kontrolüne ilişkin önemli kararlar ile yüz yüze kalmaktadırlar. Bir araştırmada, ilk önce anlamlılık düzeyinin (α) belirlenmesi gerekmektedir. Anlamlılık düzeyinin, araştırmanın doğasına uygun olarak belirlenmesi gerekirken pratikte kabul edilmiş bazı anlamlılık düzeyleri araştırmacılar tarafından kullanılmaktadır. İkinci olarak araştırmacılar, analiz edecekleri hata oranını (karşılaştırma başına veya deneysel ortak hata oranı) belirlemek durumundadırlar. Karşılaştırma başına hata oranı (α_{PC}), belirlenen α anlamlılık düzeyinde her bir karşılaştırma için yokluk hipotezinin yanlışlıkla reddedilmesi olasılığını göstermektedir. α_{PC}'nin en önemli dezavantajı, karşılaştırma sayısının (m) artması ile paralel olarak değerinin yaklaşık 1-(1-α)m kadar artmasıdır. Bu dezavantajından dolayı α_{PC} ile ilgili eleştiride bulunanlar α_{PC} yerine deneysel ortak hata oranının (α_{FW}) kontrol edilmesini önermektedirler. α_{FW} değerinin kontrolü ile hipotezler ailesindeki bir ya da daha fazla hipotezin yanlışlıkla reddedilmesi olasılığı, belirlenen α anlamlılık düzeyine ayarlanmaktadır. α_{FW} değerinin kontrolü ile ilgili işlemlerin temel avantajı, karşılaştırma sayısının artması ile paralel olarak α_{FW} değerinin artmamasıdır (4).

Tukey (5), karşılaştırma sayısına bağlı olarak 0.05'lik eşzamanlı hata oranının sırasıyla; 6 karşılaştırma için %1, 10 karşılaştırma için %0.5, 21-28 karşılaştırma için %0.2, 45-55 karşılaştırma için %0.1, 105 karşılaştırma için %0.05, 253 karşılaştırma için %0.02, 496 karşılaştırma için %0.01 ve 990 karşılaştırma için %0.005 karşılaştırma başına hata oranına karşılık geldiğini ifade etmektedir.

Bir örnek ile deneysel ortak hata kavramı daha iyi anlaşılabilir hale getirilebilir. Örneğin, dört grup olsun ve ikili olarak bu gruplara ait ortalamaların karşılaştırılması istensin. Bu durumda altı farklı ikili karşılaştırma söz konusudur. Altı farklı ikili karşılaştırmanın söz konusu olduğu bu kümeye aile (family) denmektedir. Eğer α anlamlılık seviyesinde her bir ikili karşılaştırma için t testi kullanılırsa, sadece her bir karşılaştırma için α anlamlılık seviyesinin aşılmaması garanti edilmiş olur, tüm aile için bu durum söz konusu değildir. Bundan dolayı deneysel ortak hatanın kontrol edilmesinde çoklu karşılaştırma testlerinin kullanılması gereklidir.

Bir araştırmada, araştırmacı tarafından kullanılacak hata oranı ve anlamlılık düzeyi belirlendikten sonra uygun bir, çoklu karşılaştırma testi seçilmelidir. Literatürde yer alan çoklu karşılaştırma testleri ya eşzamanlı (simultaneously) ya da adımsal (stepwise) olarak çoklu karşılaştırma işlemlerini gerçekleştirmektedir. Eşzamanlı çoklu karşılaştırma işlemlerinde tüm ikili karşılaştırmalar için sabit bir anlamlılık düzeyi dikkate alınırken, adımsal çoklu karşılaştırma testlerinde ikili karşılaştırmaların her biri için anlamlılık düzeyi yeniden belirlenmektedir.

Adımsal ve eşzamanlı çoklu karşılaştırma testleri karşılaştırıldığında deneysel ortak hata oranının kontrol edilmesinde daha etkin olması, hipotezler arasında mantıksal ilişkiler olması durumunda daha çok güce sahip olması ve yalnızca sürekli verilere değil aynı zamanda sıralı ve kategorik verilere de uygulanabilmesi gibi özelliklerinden dolayı adımsal testler öne çıkmaktadır (6).

Bu çalışmada amaç, son yıllarda daha yoğun olarak kullanılmaya başlanan ancak araştırmacılar tarafından daha az bilinen adımsal çoklu karşılaştırma testlerini tanıtmaktır.

1. Adımsal (Stepwise) Çoklu Karşılaştırma Testleri

Eşzamanlı yürütülen çoklu karşılaştırma testlerinin olması durumunda birinci tür hatanın kontrol edilmesi konusu tartışılan bir konudur. Bu kontrolün sağlanmasında, ayrı ayrı hipotezlerin test edilmesinde kullanılan önemlilik seviyesi, test edilen hipotez sayısı dikkate alınarak uyarlanmaktadır. En basit ve belki en iyi bilinen uyarlama işlemi, genel olarak kabul edilebilir birinci tür hata riskinin test edilen hipotez sayısına bölünmesidir. Bu yaklaşım Eşitlik [1] ile verilen Bonferroni eşitsizliği üzerine yapılandırılmaktadır.

$$\Pr\{\cup_{i=1}^{m} (P_i \leq \alpha/m)\} \leq \alpha \qquad [1]$$

Bonferroni eşitsizliğinde, 0 ≤ α ≤1'dir ve m (i = 1,2,…,m) test edilecek hipotez sayısını göstermektedir. Eşitlik 1 orijinal Bonferroni işlemi olarak isimlendirilir. Bonferroni işleminin iki avantajından biri, uygulanmasının kolay olması diğeri ise çok farklı çoklu test durumlarında kullanılabilir olmasıdır. İşlemin dezavantajı ise her bir hipotezi ayrı ayrı reddetme gücünün düşük olmasıdır (7).

Orijinal Bonferroni işleminin çok sayıda uyarlanmış hali geliştirilmiş ve uygulanmaktadır. Bonferroni'ye alternatif olarak uyarlanmış durumlardan altı tanesi literatürde oldukça yoğun olarak kullanılmaktadır. Bunlar sırasıyla; Holm (8), Shaffer (9), Holland ve Copenhaver (10), Hochberg (11), Hommel (12) ve Rom (13) tarafından geliştirilmiş yöntemlerdir. Bonferroni işleminin uyarlamaları olan bu yöntemlerin amacı, tümel (overall) olarak birinci tür hata riskini artırmadan ayrı ayrı testler için gücü artırmaktır.

Dunnett ve Tamhane (14) önemlilik seviyelerini düzeltmek için geliştirilen yöntemleri tek aşamalı (single step), azalan aşamalı (step down) ve artan aşamalı (step up) yöntemler olarak üç farklı şekilde kategorize etmişlerdir. Tek aşamalı işlemde (orijinal Bonferroni) olası tüm hipotezlerin ayrı ayrı testi için tek bir kriter kullanılmaktadır. Azalan ve artan aşamalı yöntemlerde ise test edilecek hipotezler kendi p

değerlerine göre sıralanır. Azalan aşamalı yöntemlerde (Holm, Shaffer, Holland ve Copenhaver yöntemleri) teste en küçük p değerine sahip hipotez ile başlanırken, artan aşamalı yöntemlerde (Hochberg, Hommel ve Rom yöntemleri) teste en büyük p değerine sahip hipotez ile başlanmaktadır. Artan aşamalı yöntemler bağımsızlık testleri için Simes (15) tarafından ispatlanan eşitsizlik üzerine oturmaktadırlar. Simes (15) tarafından, $H_0 = \{H_{(1)}, H_{(2)},\ldots,H_{(m)}\}$hipotezlerine sırasıyla karşılık gelen ve p değerlerini gösteren $p_{(1)}, p_{(2)},\ldots,p_{(m)}$ değerleri için,

$$\Pr\{_{i=1}\square^{m}(P_{(i)} \leq j\alpha/m)\} \leq \alpha \qquad (0 \leq \alpha \leq 1) \qquad [2]$$

eşitsizliğinin geçerli olduğu ispatlanarak $P_{(j)} \leq j\alpha/m$ olması durumunda j=1,2,…,m için H_0 hipotezinin reddedildiği gösterilmiştir. Bonferroni işleminin bir uyarlaması olan bu işlemde de birinci tür hata olasılığı bağımsız testler için α değerine eşittir.

Bu çalışmada, α genel olarak birinci tür hata oranını, α' test edilecek hipotezlerin her biri için ayrı ayrı kullanılacak birinci tür hata oranını ve $P_{(\mathrm{i})}$ ise $H_{(\mathrm{i})}$ i =1,2,3,…,m olmak üzere test edilecek m tane hipoteze karşılık gelen gözlenen P değerlerini gösterecektir.

Wright (16), gözlenen P değerinin P' olarak düzeltilmesini önermiştir. Wright, herhangi bir hipotez için düzeltilmiş P değerini, genel olarak kurulabilecek hipotezlerin tamamı arasında reddedilen hipotezlere ait en küçük önemlilik seviyesi olarak tanımlamıştır. m tane hipotezin orijinal Bonferroni işlemi ile test edildiği durumda $P'_{(i)} = mP_{(\mathrm{i})}$ dir. Her bir düzeltilmiş P değeri istenen α önemlilik seviyesi ile karşılaştırılmalı ve $P'_{(i)} \leq \alpha$ ise hipotez reddedilmelidir. Düzeltilmiş P değeri sabitlenmiş α yaklaşımına karşı sonuçların hangi önemlilik seviyesinde olduğunun gösterilmesini sağlar.

1.1. Orijinal Bonferroni İşlemi (Tek Aşamalı)

Orijinal Bonferroni işleminde $\alpha' = \alpha / m$ olarak hesaplanır. Her bir $H_{(\mathrm{i})}$ hipotezi $P_{(\mathrm{i})} < \alpha'$ ise reddedilir. Böylece genel olarak birinci tür hata oranı test edilecek mümkün tüm hipotezler arasında eşit olarak bölünmüş olur.

1.2. Holm İşlemi (Azalan Aşamalı)

Holm (8), reddedilen her bir hipotez için ardışık olarak düzenlenen farklı anlamlılık seviyeleri önermiştir. $H_{(1)}, H_{(2)},\ldots,H_{(m)}$ hipotezlerine karşılık gelen ve p değerlerini gösteren $P_{(1)}, P_{(2)},\ldots,P_{(m)}$ değerleri sırasıyla en büyükten en küçüğe doğru sıralansın. Holm'un önerdiği işleme göre ilk önce $p_{(m)} \geq \alpha/m$ olup olmadığına bakılır. Eğer bu koşul sağlanıyorsa diğer hipotezlere bakılmaz. Tersi durumda $H_{(m)}$ reddedilir ve sonraki hipotezin testine geçilir. Bu aşamada $P_{(m\text{-}1)} \geq \alpha / (m\text{-}1)$ olup olmadığına bakılır. Eğer bu koşul sağlanıyorsa diğer hipotezlere bakılmaz. Tersi durumda $H_{(m\text{-}1)}$ reddedilir ve sonraki hipotez olan $H_{(m\text{-}2)}$ hipotezinin test edilmesine geçilir. Süreç bu şekilde devam ettirilerek ya sonraki adıma geçilir ya da hipotezlerin test edilmesi işlemi sonlandırılır. Holm'un önerdiği ardışık artan kriter ile istatistiksel anlamlılık için istatistiksel güç kazanılır. Çünkü orijinal Bonferroni işlemi ile reddedilen herhangi bir hipotez Holm işlemi ile de reddedilmiş olacaktır.

1.3. Shaffer İşlemi (Azalan Aşamalı)

Shaffer (9), testin herhangi bir aşamasındaki doğru yokluk hipotezine ait maksimum sayının sıklıkla, Holm işleminde paydada yer alan m-i+1 (i=1,2,…,m) taneden az olduğunu göstermiştir. Shaffer tarafından önerilen işleme göre, test edilecek $H_0 : \mu_1 = \mu_2$, $H_0 : \mu_1 = \mu_3$ ve $H_0 : \mu_2 = \mu_3$ şeklindeki üç hipotezden eğer H_0: $\mu_1 = \mu_2$ hipotezi yanlış ise diğer iki hipotezden de en az biri yanlıştır. Bu durum dikkate alınarak k değeri karşılaştırılacak grup sayısını göstermek üzere k = 3,…,10 için Schaffer tarafından doğru yokluk hipotezlerinin mümkün maksimum sayılarını içeren bir tablo hazırlanmıştır. Schaffer, Holm işlemine katkıda bulunan iki öneri sunmuştur.

Shaffer'ın birinci işlemine göre; $H_0 = \{H_{(1)}, H_{(2)},\ldots,H_{(m)}$ hipotezlerine karşılık gelen ve P değerlerini gösteren $P_{(1)}, P_{(2)},\ldots,P_{(m)}$ değerleri sırasıyla en küçükten en büyüğe doğru sıralansın. ti değeri m tane hipotezden en az i-1 tanesinin yanlış olduğu varsayılarak mümkün doğru yokluk hipotezlerinin maksimum sayısı olarak tanımlanmaktadır. Eğer i^*, $p_{(i^*)} > \alpha / ti^*$ 'i sağlayan en küçük tamsayı ise $H_{(1)},\ldots,H_{(i^*\text{-}1)}$ hipotezleri reddedilir.

Shaffer'ın ikinci işlemine göre ise; ti değeri m tane hipotezden spesifik $H_{(1)},\ldots,H_{(i\text{-}1)}$ hipotezlerinin yanlış olduğu varsayılarak mümkün doğru yokluk hipotezlerinin maksimum sayısı olarak tanımlanmaktadır. Bu durumda ti değerinin birinci işlemde kullanılacak değere göre daha küçük olması söz konusudur. Shaffer'ın önerilerinden ikincisini gerçekleştirmek birincisine göre daha zor olmasına rağmen testin gücünün artmasını sağlamaktadır.

Bir örnek ile Shaffer tarafından önerilen ikinci yöntemin daha iyi anlaşılması sağlanabilir. Dört grup ve bu grupların ortalamaları arasında $\bar{X}_{\mathrm{A}} < \bar{X}_{\mathrm{B}} < \bar{X}_{\mathrm{C}} < \bar{X}_{\mathrm{D}}$ biçiminde bir sıralama söz konusu olsun. İlk karşılaştırma için $H_0 : \bar{X}_{\mathrm{A}} = \bar{X}_{\mathrm{D}}$ yokluk hipotezi kurulur. Yönteme göre mümkün doğru olabilecek maksimum sayıda H_0 aranması gerekmektedir. Buna göre; $H_0 : \bar{X}_{\mathrm{A}} = \bar{X}_{\mathrm{D}}$, $H_0: \bar{X}_{\mathrm{A}} = \bar{X}_{\mathrm{C}}$, $H_0 : \bar{X}_{\mathrm{A}} = \bar{X}_{\mathrm{B}}$, $H_0 : \bar{X}_{\mathrm{B}} = \bar{X}_{\mathrm{D}}$, $H_0 : \bar{X}_{\mathrm{B}} = \bar{X}_{\mathrm{C}}$ ve $H_0 : \bar{X}_{\mathrm{C}} = \bar{X}_{\mathrm{D}}$ hipotezleri doğru olabilir. Görüldüğü üzere birinci aşamada dikkate alınması gereken maksimum sayıdaki doğru olabilecek hipotez sayısı 6'dır. Eğer $H_0 : \bar{X}_{\mathrm{A}} = \bar{X}_{\mathrm{D}}$ hipotezi reddedilirse ya $H_0 : \bar{X}_{\mathrm{B}} = \bar{X}_{\mathrm{D}}$ ya da $H_0 : \bar{X}_{\mathrm{A}} = \bar{X}_{\mathrm{C}}$ hipotezinin test edilmesi gerekmektedir. $H_0 : \bar{X}_{\mathrm{A}} = \bar{X}_{\mathrm{C}}$ hipotezinin test edilmesi durumunda; $H_0 : \bar{X}_{\mathrm{A}} = \bar{X}_{\mathrm{C}}$, $H_0 : \bar{X}_{\mathrm{A}} = \bar{X}_{\mathrm{B}}$ ya da $H_0 : \bar{X}_{\mathrm{B}} = \bar{X}_{\mathrm{C}}$ hipotezleri doğru olabilir. $H_0 : \bar{X}_{\mathrm{A}} = \bar{X}_{\mathrm{C}}$ hipotezinin reddedilmesi durumunda $H_0 : \bar{X}_{\mathrm{B}} = \bar{X}_{\mathrm{D}}$ hipotezinin de test edilmesi gerekir. Bu durumda; $H_0 : \bar{X}_{\mathrm{B}} = \bar{X}_{\mathrm{D}}$, $H_0 : \bar{X}_{\mathrm{B}} = \bar{X}_{\mathrm{C}}$ ya da $H_0 : \bar{X}_{\mathrm{C}} = \bar{X}_{\mathrm{D}}$ hipotezleri doğru olabilir. Buradan da anlaşılacağı üzere ikinci aşamada dikkate alınacak maksimum sayıdaki doğru olabilecek hipotez sayısı 3'tür. Eğer ikinci aşama sonucunda hem $H_0 : \bar{X}_{\mathrm{A}} = \bar{X}_{\mathrm{C}}$ hem de $H_0 : \bar{X}_{\mathrm{B}} = \bar{X}_{\mathrm{D}}$ hipotezleri reddedilirse üçüncü aşamaya geçilir. Bu aşamada geriye kalan $H_0 : \bar{X}_{\mathrm{A}} = \bar{X}_{\mathrm{B}}$, $H_0 : \bar{X}_{\mathrm{B}} = \bar{X}_{\mathrm{C}}$ ya da $H_0 : \bar{X}_{\mathrm{C}} = \mathrm{X}_{\mathrm{D}}$ hipotezlerinden herhangi biri test edilebilir. Dolayısıyla üçüncü aşamada maksimum sayıda doğru olabilecek hipotez sayısı yine 3'tür. Dördüncü aşamada dikkate alınması gereken maksimum sayıda doğru olabilecek hipotez sayısının 2, beşinci aşamada ise 1 olacağı açıktır.

1.4. Holland-Copenhaver İşlemi (Azalan Aşamalı)

Holland ve Copenhaver (10), Sidak (17) tarafından geliştirilen, m karşılaştırılacak hipotez sayısını göstermek üzere α/m = 1-(1-α)1/m eşitliğini kullanarak her bir hipotez testi için bir kriter hazırlamışlardır. $H_{(1)}$, $H_{(2)},\ldots,H_{(m)}$ hipotezlerine karşılık gelen ve P değerlerini gösteren $p_{(1)}$, $p_{(2)},\ldots,p_{(m)}$ değerleri sırasıyla en küçükten en büyüğe doğru sıralansın. $P_{(i)} > 1 - (1 - \alpha)^{1/(m-i+1)}$ olacak şekilde 1 ile m arasındaki en küçük tamsayı değer i olsun. Bu durumda Holland-Copenhaver işlemine göre $H_{(m)}$'ye karşı $H_{(i)}$'ye dokunulmaksızın $H_{(i-1)}$'e karşı $H_{(1)}$ reddedilir. Holland ve Copenhaver işleminde kullanılan kriter, Holm işleminde kullanılan kriterden biraz daha geniştir, bundan dolayı ayrı ayrı her bir hipotez testi için güç biraz daha fazladır.

1.5. Hochberg İşlemi (Artan Aşamalı)

Hochberg (11), Simes (15) eşitliği üzerine oturtulmuş bir artan aşamalı yöntem geliştirmiştir. $H_{(1)}$, $H_{(2)},\ldots,H_{(m)}$ hipotezlerine karşılık gelen ve p değerlerini gösteren $P_{(1)}$, $P_{(2)},\ldots,P_{(m)}$ değerleri sırasıyla en küçükten en büyüğe doğru sıralansın. $\alpha'_{(1)}$, $\alpha'_{(2)},\ldots,\alpha'_{(m)}$ değerleri düzeltilmiş önemlilik seviyelerini göstermek üzere eğer $P_{(m)} \leq \alpha'_{(m)}$ ise tüm hipotezler reddedilir. Aksi durumda, $H_{(m)}$ alıkonularak $P_{(m-1)}$ değeri $\alpha'_{(m-1)}$ ile karşılaştırılır. Eğer $P_{(m-1)}$ değeri $\alpha'_{(m-1)}$ değerinden küçük ise geri kalan tüm hipotezler reddedilir. Eğer bu da sağlanmaz ise $H_{(m-1)}$ değeri alıkonularak $P_{(m-2)}$ değeri $\alpha'_{(m-2)}$ değeri ile karşılaştırılır ve süreç bu şekilde sürdürülür. Hochberg işlemi için, $\alpha'_{(m)} = \alpha$, $\alpha'_{(m-1)} = \alpha/2,\ldots,\alpha'_{(1)} = \alpha/m$ dir. Hipotezlerin test edilmesinde en büyük p değeri ile başlanmasının dışında Hochberg işlemi her bir hipotez için Holm işlemi ile aynı kriteri kullanmaktadır. Sonuç olarak, Hochberg işlemi Holm işlemi ile reddedilen hipotezleri benzer şekilde reddederken Holm işlemi ile incelenmeyen ancak reddedilmesi mümkün hipotezleri de test eder.

1.6. Hommel İşlemi (Artan Aşamalı)

Hommel (12) işlemi de Simes (15) eşitliği üzerine oturtulmuştur. İşlem düzeltilmiş kritik değerleri iki aşamada belirlemektedir. $J = \{i' \epsilon \{1,2,\ldots,m\}: P_{(m-i'+k)} > k\alpha/i'$; $k=1,2,\ldots,i'\}$ olsun. İşlemin ilk aşamasında elde edilen p değerleri kullanılarak J'nin eleman sayısı belirlenir. İkinci aşamada j' J'nin en büyük değeri olmak üzere $\alpha' = \alpha/j'$ kullanılarak reddetmede kullanılacak anlamlılık seviyesi elde edilir. Eğer J boş ise tüm $H_{(i)}$ (i=1,2,…,m) hipotezleri reddedilir. Eğer J boş değilse $P_{(i)} \leq \alpha/j'$ olduğu sürece $H_{(i)}$ reddedilir.

Örnek bir uygulama ile Hommel işlemi anlaşılabilir hale getirilebilir. Küçükten büyüğe doğru sıralanmış p değerleri sırasıyla $P_{(1)}$=0.0011, $P_{(2)}$=0.0145, $P_{(3)}$=0.0291 ve $P_{(4)}$=0.0562 olsun. α=0.05 anlamlılık seviyesinde test edilecek olan dört tane (m=4) hipotezin olduğu varsayılsın. Birinci aşamadaki j' değerinin belirlenmesi k değerine göre karar verilecek olan bir veya daha fazla sayıda test içerebilir ve birkaç basamak gerektirebilir. Birinci adımda i'=1 alınır ve i' her bir adım için birer birer artırılır (her bir adım içinde k=1,2,…,i').

Birinci adım yalnızca bir test içerir. m=4, i'=1, k=1 alındığında $P_{(m-i'+k)} = P_{(4-1+1)} = P_{(4)}$ = 0.0562 olur. Kritik değer ise $k\alpha/i'$=1(0.05)/1=0.05'tir. $P_{(4)}$ > 0.05 olduğundan işlem j'=1 alınarak devam ettirilir.

İkinci adım iki test içerir. İkinci adımdaki ilk test için, i'=2, k=1 alındığında $P_{(m-i'+k)} = P_{(4-2+1)} = P_{(3)}$ = 0.0291 olur. $k\alpha/i'$ = 1(0.05)/2 = 0.025'tir. $P_{(3)}$ > 0.025 olduğundan işlem devam eder. İkinci adımdaki ikinci test için i'=2, k=2 alındığında $P_{(m-i'+k)} = P_{(4-2+2)} = P_{(4)}$ = 0.0562 olur. Kritik değer ise $k\alpha/i'$ = 2(0.05)/2 = 0.05'tir. Hem $P_{(3)}$ > 0.025 hem de $P_{(4)}$ > 0.05 olduğundan işlem j'=2 alınarak devam ettirilir.

Üçüncü adım üç test içerir. Üçüncü adımdaki ilk test için i'=3, k=1 alındığında $P_{(m-i'+k)} = P_{(4-3+1)} = P_{(2)}$ = 0.0145 olur. $k\alpha/i'$ =1(0.05)/3=0.0167'dir. $P_{(2)}$ < 0.0167 olduğundan işlemin birinci aşamasına son verilir. En büyük J değeri 2 olduğundan j'=2 alınır.

Her bir test için genel olarak $P_{(m-i'+k)}$ ile elde edilen P değeri $k\alpha/i'$ ile hesaplanan kriter değer ile karşılaştırılır. Eğer $P_{(m-i'+k)} > k\alpha/i'$ ise hesaplamalar sonraki $P_{(m-i'+k)}$ ile devam eder. Bir adımdaki tüm testler için $P_{(m-i'+k)} > k\alpha/i'$ olduğu sürece j' birer birer artırılır aksi durumda hesaplamaya son verilir.

Hommel işleminin ikinci aşamasında $P_{(i)} \leq k\alpha/i'$ olduğu sürece $H_{(i)}$ reddedilir. Mevcut örnek için j'=2, α/i'=0.05/2=0.025 olduğundan $H_{(1)}$ ve $H_{(2)}$ hipotezleri reddedilir. Eğer birinci adımda $P_{(4)}$ değeri 0.05'ten küçük olsaydı tüm hipotezler reddedilirdi.

Hommel işlemi yalnızca testlerin sırasını değil aynı zamanda α' değerleri hesaplanırken, hesaplanan P değerlerini de göz önünde bulundurmaktadır.

1.7. Rom İşlemi (Artan Aşamalı)

Rom (13) istatistiksel gücü artırmak için Hochberg (11) işleminin uyarlanmış bir biçimini önermiştir. Gücün artırılması, test istatistikleri bağımsız olduğunda birinci tür hata oranının tam olarak önemsiz düzeyinin kontrol edildiği uygun düzeltilmiş anlamlılık düzeyinin belirlenmesi ile sağlanmaktadır.

Hochberg işleminde olduğu gibi $H_{(1)},H_{(2)},\ldots,H_{(m)}$ hipotezlerine karşılık gelen ve P değerlerini gösteren $P_{(1)}, P_{(2)},\ldots,P_{(m)}$ değerleri sırasıyla en küçükten en büyüğe doğru sıralansın. Eğer $P_{(m)} \leq \alpha'_{(m)}$ ise tüm hipotezler reddedilir, aksi halde $H_{(m)}$ alıkonulur ve $P_{(m-1)}$ değeri $\alpha'_{(m-1)}$ ile karşılaştırılır. İşlem sıralanmış P değeri düzeltilmiş anlamlılık düzeyinden küçük oluncaya kadar devam eder. Rom işleminde, düzeltilmiş anlamlılık düzeyinin elde edilmesi Hochberg işlemindekinden farklıdır. Her iki işlemde de $\alpha'_{(m)} = \alpha$ ve $\alpha'_{(m-1)} = \alpha/2$ eşitlikleri geçerlidir fakat geride kalan m-2 tane anlamlılık düzeyi farklıdır. Rom işleminde,

$$\alpha'_{m-i+1} = \left[\sum_{j=1}^{i-1} \alpha^j - \sum_{j=1}^{i-2} \binom{i}{j} \alpha'^{(i-j)}_{(m-j)}\right]/i \qquad [3]$$

eşitliği ile düzeltilmiş anlamlılık düzeyleri tekrarlı olarak belirlenir. Burada i = 1,2,…,m'dir. Hommel işleminde kullanılan $P_{(1)}$=0.0011, $P_{(2)}$=0.0145, $P_{(3)}$=0.0291 ve $P_{(4)}$=0.0562 değerleri kullanılarak α=0.05 anlamlılık seviyesinde Rom işlemi gösterilebilir. Buna göre;

$\alpha'_{(4)}$ = 0.05,

$\alpha'_{(3)} = \alpha'_{(4)}$ / 2 = 0.025,

$\alpha'_{(2)} = [\alpha + \alpha^2 - 3(\alpha'^2_{(3)})] / 3 = 0.0169$ ve
$\alpha'_{(1)} = [\alpha + \alpha^2 + \alpha^3 - 4(\alpha'3(3)) - 6(\alpha'^2_{(2)})] / 4 = 0.0126$ olur.

$P_{(2)} < \alpha'^2_{(2)}$ olduğundan $H_{(2)}$ hipotezi ile $P_{(2)}$ değerinden küçük P değerlerine sahip tüm hipotezler reddedilir, geri kalan hipotezler kabul edilir.

SONUÇ

Deneysel çalışmalardaki temel amaç, elde edilen sonuçlar ve denemeler arasındaki anlamlı ilişkileri açığa çıkarmaktır. Bir grup denemeden elde edilen ortalamalar arasında önemli farklılık olup olmadığının belirlenmesinde çoklu karşılaştırma testleri kullanılabilir. Uygun çoklu karşılaştırma testinin kullanılması araştırmacılara doğru, eksiksiz ve elverişli sonuçların elde edilmesinde ve elde edilen bilgiler arasındaki ilişkilerin yorumlanmasında yardımcı olur (18). Hangi durumda hangi çoklu karşılaştırma testinin kullanılacağına ya da çoklu karşılaştırma testlerinin aynı koşullar altında hangisinin daha fazla istatistiksel güce sahip olduğunun belirlenmesi amacıyla yapılmış çalışmalar az sayıda da olsa literatürde yer almaktadır (7,19,20). Jaccard ve ark. (20) tarafından yapılan çalışmada eşzamanlı çoklu karşılaştırma testleri ile ilgili hangi durumlarda hangi testin kullanılması gerektiğine ilişkin çok detaylı bilgiler verilmiştir. Olejnik ve ark. (7) tarafından yapılan çalışma ise tamamen adımsal çoklu karşılaştırma testlerini içermektedir. Bu çalışmaya göre adımsal çoklu karşılaştırma testleri içinde Holm işlemi en düşük, Rom işlemi ise en yüksek güce sahiptir. Demirhan ve ark. (19) tarafından yapılan çalışmada ise eşzamanlı çoklu karşılaştırma testlerinin hemen tamamı ile adımsal çoklu karşılaştırma testlerinin bir kısmı dikkate alınarak toplam 18 farklı çoklu karşılaştırma işlemi karşılaştırılmıştır. Bu çalışmaya göre adımsal çoklu karşılaştırma testlerinden;

- Bonferroni işleminin Varyans Analizi ile ilgili varsayımların sağlanmaması durumunda tercih edilmemesi gerektiği,
- Holm işleminin genel olarak grup sayısının az ve gruplardaki gözlem sayısının eşit olmadığı durumlarda tercih edilmesi gerektiği, Varyans Analizi ile ilgili varsayımların sağlanmamasından Holm işleminin etkilenmediği,
- Hommel işleminin gruplardaki gözlem sayılarının eşit olup olmamasından etkilendiği, gruplardaki gözlem sayılarının eşit olması durumunda tercih edilebileceği,
- Hochberg işleminin ise gerek Varyans Analizi ile ilgili varsayımların sağlanmaması durumunda gerekse gruplardaki gözlem sayılarının eşit olmaması durumlarında tercih edilmemesi gerektiği sonuçlarına ulaşılmıştır.

Adımsal çoklu karşılaştırma testleri güven aralığı belirlenmesine müsaade eden yöntemler değillerdir. Bu yöntemler eşzamanlı çoklu karşılaştırma testlerine göre deneysel ortak hata oranını kontrol altında tutarak istatistiksel gücün arttırılmasını sağlayan testlerdir (21).

KAYNAKLAR

1. Rao CV, Swarupchand U. Multiple Comparison Procedures-A Note and a Bibliography. Journal of Statistics. 2009; 16(1):66-109.
2. Games PA. 1971. Multiple Comparisons of Means. American Educational Research Journal. 1971; 8(3):531-65.
3. Hancock GR, Klockars AJ. The Quest For α: Developments in Multiple Comparison Procedures in the Quarter Century Since Games (1971). Review of Educational Research. 1996; 66(3):269-306.
4. Cribbie RA, Keselman HJ. Pairwise Multiple Comparisons: A Model Comparison Approach versus Stepwise Procedures. British Journal of Mathematical and Statistical Psychology. 2003; 56(1):167-82.
5. Tukey JW. The Philosophy of Multiple Comparisons. Statistical Science. 1991; 6(1):100-16.
6. Ludbrook J. Multiple Comparison Procedures Updated. Clinical and Experimental Pharmacology and Physiology. 1998; 25(12):1032-7.
7. Olejnik S, Li J, Supattathum S, Huberty CJ. Multiple Testing and Statistical Power with Modified Bonferroni Procedures. Journal of Educational and Behavioral Statistics. 1997; 22(4):389-406.
8. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics. 1979; 6(2):65-70.
9. Shaffer JP. Modified Sequentially Rejective Multiple Test Procedures. Journal of The American Statistical Association. 1986; 81(395):826-31.
10. Holland BS, Copenhaver MD. An Improved Sequentially Rejective Bonferroni Test Procedure. Biometrics. 1987; 43(2):417-23.
11. Hochberg Y. A Sharper Bonferroni Procedure for Multiple Tests of Significance. Biometrika. 1988; 75(4):800-2.
12. Hommel G. A Stagewise Rejective Multiple Test Procedure Based on a Modified Bonferroni Test. Biometrika.1988; 75(2):383-6.
13. Rom DR. A Sequentially Rejective Test Procedure Based on a Modified Bonferroni Equality. Biometrika. 1990; 77(3):663-5.
14. Dunnett CW, Tamhane AC. A Step-up Multiple Test Procedure. Journal of The American Statistical Association. 1992; 87(417):162-70.
15. Simes RJ. An Improved Bonferroni Procedure for Multiple Tests of Significance. Biometrika. 1986; 73(3):751-4.
16. Wright SP. Adjusted P-Values for Simultaneous Inference. Biometrics. 1992; 48(4):1005-13.
17. Sidak Z. Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. Journal of The American Statistical Association. 1967; 62(318):626-33.
18. Lowry SR. Use and Misuse of Multiple Comparisons in Animal Experiments. Journal of Animal Science. 1992; 70(6):1971-7.

19. Demirhan H, Dolgun NA, Demirhan YP, Dolgun MÖ. Performance of Some Multiple Comparison Tests Under Heteroscedasticity and Dependency. Journal of Statistical Computation and Simulation. 2010; 80(10):1083-100.
20. Jaccard J, Becker MA, Wood G. Pairwise Multiple Comparison Procedures: A Review. Psychological Bulletin. 1984; 96(3):589-96.
21. Hancock GR, Klockars AJ. Finite Intersection Tests: A Paradigm for Optimizing Simultaneous and Sequential Inference. Journal of Educational and Behavioral Statistics. 1997; 22(3):291-307.