ARDIŞIK ÇOKLU KARŞILAŞTIRMA TESTLERİNDEN FDR TESTİ ÜZERİNE BİR SİMÜLASYON UYGULAMASI

Aybüke KOCA

DANIŞMAN
Yrd. Doç. Dr. İbrahim KILIÇ
İSTATİSTİK

Eylül 2013

AFYON KOCATEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

ARDIŞIK ÇOKLU KARŞILAŞTIRMA TESTLERİNDEN FDR TESTİ ÜZERİNE BİR SİMÜLASYON UYGULAMASI

Aybüke KOCA

DANIŞMAN
Yrd. Doç. Dr. İbrahim KILIÇ

İSTATİSTİK

Eylül 2013

TEZ ONAY SAYFASI

Aybüke Koca tarafından hazırlanan "**Ardışık Çoklu Karşılaştırma Testlerinden FDR testi Üzerine bir Simülasyon Uygulaması**" adlı tez çalışması lisansüstü eğitim ve öğretim yönetmeliğinin ilgili maddeleri uyarınca ..../......../........ tarihinde aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü **İstatistik Anabilim Dalı'nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman **:** Yrd. Doç. Dr. İbrahim KILIÇ

Başkan **:** Prof. Dr. İsmet DOĞAN

Afyon Kocatepe Ü. Tıp Fakültesi

Üye **:** (Danışman) Yrd. Doç. Dr. İbrahim KILIÇ

Afyon Kocatepe Ü. Veteriner Fakültesi

Üye **:** Yrd. Doç. Dr. Sinan SARAÇLI

Afyon Kocatepe Ü. Fen-Edebiyat Fakültesi

Afyon Kocatepe Üniversitesi
Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
........./......../........ tarih ve
………………. sayılı kararıyla onaylanmıştır.
……………………………….
Prof. Dr. Mevlüt DOĞAN
Enstitü Müdürü

BİLİMSEL ETİK BİLDİRİM SAYFASI

Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,

- Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,

- Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,

- Atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,

- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,

- Ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

08 / Temmuz / 2013

Aybüke KOCA

ÖZET

Yüksek Lisans Tezi

ARDIŞIK ÇOKLU KARŞILAŞTIRMA TESTLERİNDEN

FDR TESTİ ÜZERİNE BİR SİMÜLASYON UYGULAMASI

Aybüke KOCA

Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü

İstatistik Anabilim Dalı

Danışman: Yrd. Doç. Dr. İbrahim KILIÇ

Varyans Analizi ikiden fazla gruplar söz konusu olduğunda gruplar arasında farklılık olup olmadığını belirlemektedir. Fakat farklılığın hangi gruptan kaynaklandığı belirleyebilmek için çoklu karşılaştırma testlerinden yararlanılır. Çoklu karşılaştırma için birçok teknik söz konusu olduğundan, hangi tekniğin kullanılacağı konusunda göz önünde bulundurulması gereken en önemli noktalardan biri, karşılaştırılacak grup sayısı, diğeri ise veri setinin taşıdığı özelliklerdir. İlgilenilen yığını oluşturan birimler için göz önünde bulundurulan değişkenin ölçüm düzeyi, gruplardan herhangi birisinin diğerlerine göre referans bir grup olup olmayacağı da araştırmacının göz önünde bulundurması gereken önemli bir konudur.

Bu çalışmada varyans analizi sonucunda sıfır ön savının reddedilmesi durumunda gruplar arasındaki farkın belirlenmesinde son zamanlarda daha çok kullanılan karşılaştırma testleri incelenmiştir. Çalışmada sırası ile konuya ilişkin literatür bilgileri ve daha önce yapılan çalışmalar özetlenmiş, daha sonra varyans analizi ve sıfır ön savının reddi durumunda kullanılabilecek çoklu aralık, çoklu karşılaştırma ve ardışık çoklu karşılaştırma testlerine yer verilmiştir. Çalışmanın uygulama kısmında MATLAB paket programı yardımıyla simülatif olarak farklı grup sayıları için farklı örneklem hacimlerinde veriler türetilerek

FDR testi üzerine yapılan bir uygulamaya yer verilmiş, elde edilen bulgular ilgili çizelgelerde verilmiştir.

2013, xi + 56 sayfa

Anahtar Kelimeler: Ardışık Çoklu Karşılaştırma, FDR, Simülasyon

ABSTRACT
M.Sc. Thesis

A SIMULATION APPLICATION ON FDR TEST, A SEQUANTIAL MULTIPLE COMPARISON TESTS

Aybüke KOCA

Afyon Kocatepe University
Graduate School of Natural and Applied Sciences
Department of Statistics
Supervisor: Assistant Professor Dr. İbrahim KILIÇ

Analysis of Variance, in the case of more than two groups determine whether there were differences between the groups. However, due to the difference in order to determine which group benefited from multiple comparison tests. Because there are many techniques for multiple comparisons, two of the most important points to decide which technique is the most appropriate are the numbers of the groups and the features of the data set. Also it's important to consider for the researcher that whether any of the group is reference for the others and the measurement scale of the variable which is considered for the units of the mass.

In this study, multiple comparison tests which are recently more widely used to determine the differences between groups when the null hypothesis is rejected are examined. In the study some of the early studies summarized after some knowledge about the literature is given, than some information about analysis of variance and multiple comparison, multiple difference and sequential multiple comparison tests which can be used when the null hypothesis is rejected, are given respectively. In application part of the study MATLAB software is used to obtain the date sets simulatively for different sample sizes and for the different numbers of variables for the application on FDR test. The results of the analysis are given in the related tables.

2013, xi + 56 pages

Key Words: Multiple Comparison, FDR, Simulation

TEŞEKKÜR

Bu çalışmanın gerçekleşmesindeki katkılarından dolayı değerli danışmanım Yrd. Doç. Dr. İbrahim KILIÇ'a, tezimi bitirmek için destek veren her türlü olanağı sağlayan yardımlarını esirgemeyen hocam Yrd. Doç. Dr. Sinan SARAÇLI'ya, tez çalışması ile ilgili bütün kaynaklara ulaşmamda yardımcı olan Prof. Dr. İsmet DOĞAN'a, bütün bölüm hocalarıma, bu çalışmada bana her daim yardımcı olan tüm arkadaşlarıma ve Beni bugünlere getiren desteğini eksik etmeyen canım aileme,

Ayrıca konunun belirlenmesinde Yrd. Doç. Dr. Nurhan DOĞAN ve Prof. Dr. İsmet DOĞAN tarafından yapılmış olan **"Birinci Tür Hata'nın Kontrolünde Yanlış Bulgu Oranı (False Dıscovery Rate) Yaklaşımı**" çalışma belirleyici olduğu için yazarlara sonsuz TEŞEKKÜR'lerimi sunuyorum…

Aybüke KOCA

AFYONKARAHİSAR, 2013

İÇİNDEKİLER

Sayfa no

SİMGELER ve KISALTMALAR DİZİNİ

Simgeler

α	Anlam Seviyesi
α_{FW}	Deneysel ortak hata oranı
α_{PC}	Karşılaştırma başına düşen hata oranı
H_0	Yokluk hipotezi
K	Grup sayısı
M	Yokluk hipotez sayısı
m_0	Doğru yokluk hipotez sayısı
μ	Ana kitlenin ortalaması
n_i	i. Örneklemdeki gözlem sayısı
P	Büyüklük sırasına dizilmiş ortalamalar arası farkın kademe sayısı
p_i	Yokluk hipotezine karşılık gelen p değerleri
S_i^2	i.gruba ait varyans
V	Hata serbestlik derecesi (n-k)
W	İkinci tip hataya karşı yapılan birinci tip hata
x_{ij}	i. örneklemdeki j. gözlem değeri
$\bar{x}$	Gözlem değerlerinin ortalaması
σ^2	Ana kitlenin varyansı

Kısaltmalar

ANOVA	Analysis of Variance, Varyans Analizi
EKK	En Küçük Kareler Metodu
FDR	False Discovery Rate, Yanlış Sınıflama Oranı
FWER	Familywise Error Rate
GA	Güven Aralığı
GAKT	Gruplar Arası Kareler Toplamı
GİKT	Gruplar İçi Kareler Toplamı
HKT	Hata Kareler Toplamı
HSD	Honestly Significant Difference, En Güvenilir Anlamlı Fark
LS	Least Squares, En Küçük Kareler Metodu
LSD	Least Significant Diference, En Küçük Anlamlı Fark Testi
MCA	All-pairwise Multiple Comparison, Tüm İkili Çoklu Karşılaştırmalar
MCB	Multiple Comparison with the Best, En İyi ile Çoklu Karşılaştırmalar
MCC	Multiple Comparison with a Control, Bir Kontrol ile Çoklu Karşılaştırmalar
MCM	Multiple Comparison with the Mean, Ortalama ile Çoklu Karşılaştırmalar
PCER	Per-Comparison Error Rate
PFDR	Positive False Discovery Rate, Pozitif Yanlış Bulgu Oranı
R-E-G-W-Q	Ryan-Einot-Gabriel-Welsch Q range test
R-E-G-W-F	Ryan-Einot-Gabriel-Welsch F range test
SD	Serbestlik Derecesi
S-N-K	Student Newman Keuls
WSD	Wholly Significant Difference

ŞEKİLLER DİZİNİ

Sayfa

ÇİZELGELER DİZİNİ

Sayfa

1.GİRİŞ

Varyans Analizi (Değişke Çözümlemesi), ikiden fazla grup ortalaması arasındaki farkın anlamlılığının sınanmasında kullanılan parametrik bir istatistiksel analiz yöntemidir. Varyans Analizinde test edilecek değişkenlerin sürekli sayısal değişken olmaları mutlak şarttır.

İlk olarak R.A. Fisher tarafından 1918-1924 yıllarında öne sürülen Varyans Analizi çözümlemesi kısa zamanda çok geniş bir kullanım alanı bulmuştur (Çömlekçi 2003).

Bu çalışmada öncelikle tek faktörlü Varyans Analizi ve uygulaması sonucunda F testinden yararlanılarak ikiden fazla grup ortalamaları arasındaki farkların anlamlı olup olmadığının tespit edilmesinde kullanılan testlerin özelliklerinden bahsedilmiştir. Bu konuda daha önce yapılmış çalışmalar şöyle özetlenebilir;

Tukey (1949) çalışmasında; Varyans Analizi uygulaması sonunda grup ortalamaları arasındaki farkı belirlemek amacıyla basit ve kolay anlaşılır bir yöntem önermiş ve örnekle açıklamıştır.

Keselman ve Rogen (1978) çalışmalarında; Tukey metodunun geliştirilmişi olan Kramer (1956), Spiqtvoll ve Stoline (1975), Hochberg (1976), Games ve Howell (1976) ve Scheffe (1959) yöntemlerini normal ve çarpık dağılımlar örnek alındığında birinci tip hata oranlarını, varyans homojenliğini ve dengesiz örnek büyüklüğü durumlarındaki duyarlılığını karşılaştırmışlardır. Bunun sonucunda sadece Games ve Howell (1976) ve Hochberg (1976) yöntemlerinin belirli anlam seviyesinde birinci tip hata oranını kontrol ettiğini ortaya koymuşlardır.

Holm (1979) çalışmasında; basit ve oldukça geniş kullanım alanına sahip ardışık çoklu karşılaştırma yöntemi sunmuştur. Bonferroni yönteminin geliştirilmişi olan bu yöntemle doğru hipotezlerin bütün kombinasyonları için ayrı ayrı belirlenmiş anlam seviyesinde birinci tip hatanın korunduğunu göstermiştir.

Simes (1986) çalışmasında; her bir testin sıralı p değerlerini temel alan geliştirilmiş bir yöntem sunmuştur. Bu yöntem klasik Bonferroni yöntemine göre daha güçlü fakat uygulaması daha kolaydır. Yöntemde birinci tip hata olasılığının çok değişkenli normal ve gamma test istatistikleri için α anlam seviyesini aşmadığını, bağımsız testler içinse α'ya eşit olduğunu yaptığı simülasyon çalışmalarıyla göstermiştir. Özellikle yüksek korelasyonlu gözlem değerleri içerdiğinde Bonferroni yönteminden daha avantajlı olduğunu öne sürmüştür.

Shaffer (1986) çalışmasında; Holm yönteminden daha güçlü geliştirilmiş bir Bonferroni yöntemi sunmuştur. Testin herhangi bir aşamasındaki doğru yokluk hipotezine ait maksimum sayının Holm yöntemindekinden daha az olduğunu göstermiştir. Aynı zamanda doğru yokluk hipotezlerinin mümkün maksimum sayılarını içeren bir tablo hazırlamıştır. Yöntemin bütün problemlerde kolaylıkla kullanılabileceğini örneklerle göstermiştir.

Holland ve Copenhaver (1987) çalışmalarında; ardışık çoklu test yöntemlerinde önemli durumlarda karşılaşılan varsayımları belirlemiş ve gerekli olan anlam seviyesi uygun olduğunda diğerlerinin yerine kullanılabilecek yeni bir yöntem önermiştir. Yöntemin Holm işleminden biraz daha geniş kriterle hipotez testleri için daha güçlü olduğunu göstererek ve sayısal örneklerle açıklamıştır.

Hochberg (1988) çalışmasında; Holm yönteminden daha anlaşılır, p değerlerini ayrı ayrı göz önünde bulunduran basit bir yöntem önermiştir. Holm yönteminde kritik değere eşit ya da daha küçük p değerli sadece bir hipotez reddedilirken, yeni yöntemle kritik değere eşit ve küçük olan bütün hipotezlerin reddedildiğini göstermiştir.

Hommel (1988) çalışmasında; hipotezlerin bütün kombinasyonlarını göz önünde bulunduran ardışık çoklu karşılaştırma yöntemi sunmuştur. Hommel yöntemi düzeltilmiş α değerlerini hesaplarken yalnızca testlerin sırasını değil aynı zamanda hesaplanan p değerlerini de dikkate almıştır. Temeli kapalı test prosedürüne dayanan bu yöntemle çoklu α seviyesinin de kontrol altına aldığını göstermiştir.

Wright (1992) çalışmasında; her bir hipotezin düzeltilmiş p değerinin, seçilen α değerinden az olduğunda, hipotezin α'dan daha küçük deneysel hata oranı ile reddedileceğini göstermiş ve simülasyon sonuçlarını ortaya koymuştur. Varyans Analizindeki çoklu karşılaştırma testlerinde kullanılan düzeltilmiş p değeri ile Bonferroni yöntemini temel alan düzeltilmiş p değerlerini ve bu yöntemin Holm, Hochberg, ve Hommel tarafından geliştirilmiş hallerini örneklerle sunmuştur. Geliştirilmiş Bonferroni yönteminin daha güçlü olduğunu göstermiştir.

Dunnett ve Tamhane (1995) çalışmalarında; artan aşamalı çoklu karşılaştırma yöntemlerini önermişlerdir. Yöntemin en az anlamlı test istatistiğine karşılık gelen hipotezle başlayıp en önemlisine doğru devam ettiğini ve ilk anlamlı test sonucu elde edildiğinde bitirilerek kalan hipotezlerin reddedildiğini göstermişlerdir. Aynı zamanda artan aşamalı ve azalan aşamalı farklı test yöntemlerini de karşılaştırmışlardır.

Benjamini ve Hochberg (1995) çalışmalarında; çoklu anlamlılık testlerinde en yaygın yaklaşım olan deneysel hata oran kontrolünden farklı yeni bir yaklaşım sunmuşlardır. Bu yöntem yanlışlıkla reddedilen hipotezlerin beklenen oranlarının kontrolüdür ve yanlış bulgu oranı (FDR) olarak adlandırmışlardır. Tüm hipotezler doğru olduğunda bu hata oranının deneysel ortak hata oranına eşit olduğunu göstermişlerdir. Basit Bonferroni yönteminde bağımsız test istatistikleri için yanlış bulgu oranını kontrolünü ispatlamış ve güçte önemli bir artış sağladığını simülasyon çalışması ile göstermişlerdir.

Benjamini ve Hochberg (2000) çalışmalarında; çoklu anlamlılık testlerinde yanlış reddedilen hipotezlerin beklenen oranlarını kontrol eden ve FDR olarak adlandırdıkları yeni bir yaklaşım getirmişlerdir. Bu yöntem bağımsız test istatistikleri için FDR'nin kontrol edildiğini göstermiştir. İlk olarak doğru hipotez sayısını Hochberg ve Benjamini (1990)' da bahsedildiği gibi tahmin edildiğinde ve bu tahmin Benjamini ve Hocahberg (1995)' deki yöntemde kullanıldığında uygulanabilecek bir yöntem sunmuşlardır. Bu yöntemde bağımsız test istatistiklerin simülasyon çalışması ile FDR'nin kontrol edildiğini ve deneysel ortak hata oranını kontrol eden metodlardan daha güçlü olduğunu göstermişlerdir.

Benjamini ve Yekutieli (2001) çalışmalarında; Benjamini ve Hochberg önerileri üzerine bu benzer yöntemle doğru yokluk hipotezine uygun her bir pozitif bağımlı test istatistiğine sahip olduğunda aynı zamanda FDR'yi de kontrol ettiğini ispatlamışlardır. Pozitif korelasyonlu çok değişkenli normal ve t test istatistiklerinin ve basit kontrollü birçok uygulamaların karşılaştırmalarını içeren uygulama ile bütün problemleri kapsayacak şekilde genellemişlerdir. FDR yöntemini bağımlılığın diğer bütün çeşitleri için tutarlı hale getirmişlerdir. Böylece önerilen FDR kontrolü için problem çeşitliliğini büyük ölçüde artırmışlardır.

Bu araştırmada, Varyans Analizi sonucunda kullanılan eş zamanlı ve ardışık çoklu karşılaştırma testlerine alternatif bir teknik olarak FDR testinin performansının incelenmesi amaçlanmıştır.

2. LİTERATÜR BİLGİSİ

2.1. Varyans Analizi

Yirminci yüzyılın en büyük istatistikçilerinden olan Ronald Aylmer Fisher tarafından geliştirilen Varyans Analizi (Analysis of Variance- ANOVA) birçok kaynağa katkı sağlamıştır (Gurarie 2008). Varyans analizi (ANOVA), gruplar arasındaki farklılıkları karşılaştırmaya olanak tanıyan istatistiksel süreçler bütünüdür (İngersoll 2010).

Varyans Analizi, parametrik test varsayımları yerine getirildiğinde ölçümle belirtilen ana kitlelerde normal dağılım gösteren iki ve ya daha fazla sayıda grup ortalaması arasında fark olup olmadığını belirlemektedir. Aynı zamanda bu farkın önemini ve farkı oluşturan nedenleri kontrol etmek için de kullanılır (İrgüren 2001).

Bağımsız örneklemler için Varyans Analizi ise; tek bir bağımsız değişkene ilişkin iki veya daha fazla grubun bağımlı bir değişkene göre ortalamalarının karşılaştırılarak, ortalamalar arası farkın belirli bir güven düzeyinde anlamlı olup olmadığını test etmek için kullanılan bir yöntemdir (Ural ve Kılıç 2011).

Çizelge2.1 Tek Yönlü Varyans Analizi Veri Düzeni(Polat 2006)

Anakütle	1	2	….	K	
Örneklem hacmi	n_1	n_2	….	n_k	
Örneklem birimleri	X_{11} X_{12} . . . $X_1 n_1$	X_{21} X_{22} . . . $X_2 n_2$	. . . . . .	X_{k1} X_{k2} . . . $X_k n_k$	
Örneklem birimleri toplamı	X_1	X_2	….	X_k	X
Örneklem ortalaması	$\bar{X}_1$	$\bar{X}_2$	….	$\bar{X}_k$	$\bar{X}$

Varyans Analizi bazı temel varsayımlara dayanır. Bu varsayımlar aşağıda açıklanmıştır.

1. varsayım: Hata terimleri 0 ortalama ve σ^2 varyans ile normal dağılıma sahiptir.
2. varsayım: Hata terimlerinin varyansları homojendir.
3. varsayım: Hata terimleri birbirinden bağımsızdır.

ANOVA modelindeki parametrelerin En Küçük Kareler (EKK, LS) tahmin edicileri normal dağılım varsayımı altında en etkin tahmin edicilerdir. Dolayısıyla LS tahmin edicilerine dayanan F testide en güçlü testtir. ANOVA da güvenilir ve etkin sonuçlar elde etmek için normallik varsayımının sağlanması gerekir. Aksi takdirde parametre tahmin edicilerinin ve buna bağlı olarak F test istatistiklerin güvenilirliği de düşük olur (Şenoğlu ve Acıtaş 2011).

Bilindiği üzere, varyans analizinde kurulan hipotez;
H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ ve
H_1: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \ldots \neq \mu_k$ ya da en az bir ortalama farklı şeklindedir.

Araştırmacı, Varyans Analizi sonucu H_0 hipotezini reddedip, H_1 hipotezini kabul etmesi durumunda, farklılığın hangi grup ya da gruplardan kaynaklandığını tespit etmek üzere çoklu karşılaştırma testi seçimi yapmak durumundadır. Ancak, karşılaştırma testi seçiminde isabetli istatistik türünün seçimi, hipotezlerin I. ve II. tip hata risklerini asgari seviyeye indirme yönünde oldukça önem taşımaktadır (Kayri 2009).

Çizelge 2.2 Tek Yönlü Varyans Analizi Tablosu(Polat 2006)

Değişim kaynağı	**Kareler Toplamı (KT)**	**Serbestlik Derecesi (SD)**	**Kareler Ortalaması (KO)**	**F istatistiği**
Gruplar arası	GAKT	k-1	GAKO=GAKT/(k-1)	$\frac{\text{GAKO}}{\text{GİKO}}$
Gruplar içi	GİKT=HKT	n-k	GİKO=GİKT/(n-k)	
Genel	GKT	n-1	GKO=GKT/(n-1)	

Burada toplam değişkenlik ikiye ayrılır. Birincisi her bir ana kütleden bağımsız olarak çekilen örneklemlerin kendi içindeki değişkenlik, diğeri ise k örneklem grubu arasındaki değişkenliktir. Genel kareler toplamı;

$$\mathrm{GKT} = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^{2} - \left[\frac{\sum_{i=1}^{k} X_i^2}{n}\right] \qquad (2.1)$$

Gruplar arası kareler toplamı;

$$\mathrm{GAKT} = \sum_{j=1}^{n}\left[\frac{\left(\sum_{j=1}^{n} x_j\right)^2}{n_j}\right]-\left[\frac{\sum_{i=1}^{k} X_i^2}{n}\right] \qquad (2.2)$$

Gruplar içi kareler toplamı (hata kareler toplamı);

$$\mathrm{HKT} = \mathrm{GKT} - \mathrm{GAKT} \qquad (2.3)$$

$\frac{GAKO}{HKO}$ oranı payı (k-1), paydası (n-k) serbestlik dereceli F dağılımına sahiptir

Hesaplanan değer, tablodan $F_{(\mathrm{k}-1),(\mathrm{n}-\mathrm{k}),\alpha}$ tablo değeri ile karşılaştırılır.

$$\frac{GAKO}{HKO} \le F_{(\mathrm{k}-1),(\mathrm{n}-\mathrm{k}),\alpha} \qquad (2.4)$$

ise H_0 kabul edilir ve grup ortalamaları arasında fark olmadığına karar verilir.

$$\frac{GAKO}{HKO} \ge F_{(\mathrm{k}-1),(\mathrm{n}-\mathrm{k}),\alpha} \qquad (2.5)$$

ise H_0 reddedilir ve grup ortalamaları arasında fark olduğuna karar verilir (Şenoğlu ve Acıtaş 2011).

Varyans analizinde;

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$

Şeklinde ifade edilen sıfır hipotezinin reddedilmesi durumunda grup ortalamaları arasındaki farklılığın hangi grup ve ya gruplardan kaynaklandığını belirlemek amacıyla literatürde yaygın olarak kullanılan ikili ve çoklu karşılaştırma metotları tanıtılmıştır.

Varyans Analizi sonucunda yapılması gereken ilk iş, grupların varyanslarının eşit olup olmadığına yani incelenen grupların aynı ana kütleden ya da aynı dağılıma sahip bir ana kütleden çekilip çekilmediğine karar vermektir.

Varyansların homojen olması durumunda araştırmacı, çoklu karşılaştırma veya çoklu aralık testlerinden birini tercih etmek durumundadırlar. Uygulanan Varyans Analizi sonucunda sıfır hipotezinin red edilmesi durumunda hangi gruplar arasında istatistiksel olarak anlamlı bir fark olduğunu belirleyebilmek için kullanılabilecek çoklu karşılaştırma ve çoklu aralık testleri aşağıda verilmektedir.

2.2 Çoklu Karşılaştırma Testleri

Çoklu karşılaştırma testleri, Varyans Analizi sonucunda ilgilenilen gruplar arasındaki farklılığın anlamlılığını test emek amacıyla kullanılır. Çoklu karşılaştırma testi sonrasında alt gruplar hakkında ayrıntılı bilgi elde edilebilir. Burada dikkat edilmesi gereken iki önemli nokta vardır. Bunlar Varyans Analizi uygulayabilmenin varsayımları ve karşılaştırılacak grup sayısının fazlalığıdır. Varyans Analizi varsayımları bilindiği üzere gözlemlerin normal dağılış sergilemesi gerekliliği, grup varyanslarının homojen olması ve grupların birbirinden bağımsız olmasıdır. Bu varsayımlar çoklu karşılaştırma testlerinin yapılabilmesi için de sağlanmalıdır. Varsayımlara da bağlı olarak grup hacimlerinin eşit olması çoklu karşılaştırma testlerinin performansını da etkiler. Hangi çoklu karşılaştırma testinin kullanılacağına karar verilmesi kolay bir işlem değildir. Bazı çoklu karşılaştırma testleri sağlanan ve ihmal edilen varsayımlara göre daha etkili sonuçlar vermektedir. Bu konu hakkında birçok çalışma ve bilgi olmasına karşın kendi içerisinde oldukça karışık bir süreçtir (Demirhan ve diğerleri 2011).

Rao ve Swarupchand (2009)' a göre çoklu karşılaştırma terimi bir grup içindeki varyanslar, oranlar ya da ortalamalar arasındaki farklılığın istatistiksel anlamlılığı için yapılan testleri ifade etmektedir. Çoklu karşılaştırma işlemleri, çokluluk etkisinden kaynaklanan hatalı sonuç çıkarımların düzenli kontrolünü göz önünde bulunduran istatistiksel işlemlerdir. Çoklu karşılaştırma işlemleri uygulamalardaki önemliliğinden dolayı temel bir problemdir ve farklı yollarla kullanıcılara yol göstermektedirler. Araştırmacıların amacına göre literatürde görülen dört tip çoklu karşılaştırma işlemi söz konusudur (Doğan ve Doğan 2013). Bunlar;

a- Tüm ikili çoklu karşılaştırmalar (all-pairwise multiple comparison (MCA), $i \neq j$ olmak üzere tüm μ_i-μ_j farklarının karşılaştırılması dikkate alınmaktadır),

b- En iyi ile çoklu karşılaştırmalar (multiple comparison with the best (MCB), $i \neq j$ ve i=1, 2, ……, k olmak üzere tüm μ_i-$max \mu_j$ farklarının karşılaştırılması dikkate alınmaktadır),

c- Bir kontrol ile çoklu karşılaştırmalar (multiple comparison with a control (MCC), i=1, 2,…, k-1 olmak üzere tüm μ_i-μ_k farklarının karşılaştırılması dikkate alınmaktadır),

d- Ortalama ile çoklu karşılaştırmalar (multiple comparison with the mean (MCM), i=1, 2,……, k olmak üzere tüm μ_i-$\bar{\mu}$ farklarının karşılaştırılması dikkate alınmaktadır).

Araştırmada, farklılık yaratan grup ya da grupları tespit etmek üzere birçok çoklu karşılaştırma testi bulunmakla birlikte, bunların doğru bir şekilde seçimi bazı varsayımlar gerektirmektedir. Çoklu karşılaştırma testlerine ait istatistik türlerinin seçiminde, önemli unsurlardan olan gruplar arası varyansın eşit olup-olmama özelliği önem taşımaktadır.

Varyansların homojen olması durumunda kullanılacak karşılaştırma testleri genel itibariyle iki gruba ayrılmaktadır.

Çoklu Aralık Testleri: Grup ortalamalarına ilişkin homojen alt setler (homogeneous subset) oluşturarak, gruplardan farklı olanları tespit etmeye çalışmaktadır.

Çoklu Karşılaştırma Testleri: Her grubu sırasıyla diğer gruplarla teker teker kıyaslayarak ve bir karşılaştırma matrisi ile gruplardan farklı olanları tespit etmeye çalışmaktadır.

2.2.1.Varyanslar Homojen Olduğunda Kullanılacak Çoklu Karşılaştırma ve Aralık Testleri

Varyanslar homojen olduğunda kullanılan hem çoklu karşılaştırma ve hem de aralık testleri grubundan Tukey HSD testi tüm ikili karşılaştırmaları yapmakta ve aynı zamanda student t testini kullanmaktadır. Tukey testinin büyük hacimli örneklemlerde Bonferroni testinden, eşit örneklem büyüklüklerinde ise Hochberg GT2 testinden daha güçlü olduğu ifade edilmektedir. Hochberg GT2 testi Tukey testine çok benzerdir fakat student maksimum modulus tablo değerini kullanmaktadır. Tukey Kramer testi örneklem büyüklükleri eşit olmadığı durumlarda kullanılır ve genellikle Hochberg GT2 testinden daha güçlüdür. Gabriel testi eşit örneklem büyüklüklerinde Hocberg GT2 testine benzerdir fakat farklılık gösterdiğinde daha güçlü olduğu ifade edilmektedir. Scheffe anlamlılık düzeyi test edilen grup ortalamalarının tüm dik doğrusal kombinasyonlarına olanak sağlamaktadır. Genellikle diğer testlere oranla daha güçlüdür (SPSS 2011).

Bahsedilen testlerin formülleri ve karar kriterleri Çizelge 2.3 ‘de verilmektedir (Carmer and Walker 1985, SPSS 2011).

Çizelge 2.3 Varyanslar Homojen Olduğunda Kullanılan Çoklu Karşılaştırma ve Aralık Testleri

TEST	KRİTİK DEĞER	KARAR
Tukey HSD ve Tukey Kramer	$HSD=\begin{cases} q_{\alpha,k,v}\sqrt{\frac{HKO}{n}} & n_1=n_2=\cdots=n_k=n \\ q_{\alpha,k,v}\sqrt{\frac{HKO}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} & n_1\neq n_2\neq\cdots\neq n_k \end{cases}$ $v=(n-k)$	$\lvert\bar{X}_i-\bar{X}_j\rvert > HSD$
Hochberg GT2	G.A=$(\bar{X}_i-\bar{X}_j)\pm\ m_{\alpha,k^*,v}\sqrt{HKO\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$ $k^*=k(k-1)/2$ $v=(n-k)$	G.A '0' değerini içermiyor ise
Gabriel	G= $\sqrt{HKO}\left(\frac{1}{\sqrt{2n_i}}+\frac{1}{\sqrt{2n_j}}\right)m_{\alpha,k^*,v}$ $k^*=k(k-1)/2$ $v=(n-k)$	$\lvert\bar{X}_i-\bar{X}_j\rvert >$ G
Scheffe	G.A=$(\bar{X}_i-\bar{X}_j)\pm\ \sqrt{(k-1)F_{\alpha,(k-1),v}}\sqrt{HKO\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$ $v=(n-k)$	G.A '0' değerini içermiyor ise

Burada;

$q_{\alpha,k,v}$: Tukey ve SNK tablo değerini

$m_{\alpha,k^*,v}$: Studentized maksimum modulus tablo değerini

$F_{\alpha,(k-1),v}$: F tablo değerini göstermektedir.

Çizelge 2.3'te belirtilen karar kriterleri sağlandığı takdirde yokluk hipotezi reddedilerek gruplar arasında farklılık olduğuna karar verilir.

2.2.2. Varyanslar Homojen Olduğunda Kullanılan Çoklu Aralık Testleri

Varyanslar homojen olduğunda kullanılan çoklu aralık testlerinden SNK testi student t istatistiğini kullanmaktadır. İkili grup karşılaştırmalarında tek bir anlamlılık düzeyi yerine her aşamada değişen bir anlamlılık düzeyini tercih etmektedir. Duncan testi Tukey HSD ve SNK'dan daha güçlüdür fakat birinci tip hatayı kontrol etmekte yetersiz kalmaktadır. Testin tüm aşamalarında kullanılan anlamlılık düzeyi bakımından SNK'dan farklılık göstermektedir (Atil ve Ünver 2011).

Tukey WSD testi SNK'ya benzer bir mantıkla yürütülmekte olup, diğer testlerde olduğu gibi grup örneklem büyüklüğünün eşit olmadığı durumda birinci tip hatayı kontrol edememektedir. Benzer şekilde R-E-G-W-F ve R-E-G-W-Q testleri de çoklu aralık testi olup, R-E-Q-W-Q istatistiği, Duncan testinde olduğu gibi α anlamlılık düzeyini grup sayısına göre esnek kılabilmektedir. Waller Duncan testi ise Bayesian bir yaklaşım sunan ve örneklem sayıları eşit olmadığında Bancroft'un önerdiği grup sayılarının harmonik değerini kullanan bir test istatistiğidir. Ayrıca birinci ve ikinci tip hatalara karşı koruma düzeyi yüksek bir yaklaşım sunmaktadır (Kayri 2009).

Bahsedilen çoklu aralık testleri Çizelge 2.4'de gösterilmiştir (Carmer and Walker 1985, SPSS 2011).

Çizelge 2.4 Varyanslar Homojen Olduğunda Kullanılan Çoklu Aralık Testleri

TEST	KRİTİK DEĞER	KARAR
Tukey WSD	F = $\frac{n(\overline{X_i}-\overline{X_j})}{2\,HKO}$, q = $\sqrt{2F}$, $n_1 = n_2 = \cdots = n_k = n$ $v = (n-k)$	q > $q_{\alpha,k,v}$
SNK	$W_p = \begin{cases} q_{\alpha,p,v}\sqrt{\frac{HKO}{n}}, & n_1 = n_2 = \cdots = n_k = n \\ q_{\alpha,p,v}\sqrt{\frac{HKO}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}, & n_1 \neq n_2 \neq \cdots \neq n_k \end{cases}$ $v = (n-k)$	$\lvert\overline{X_i}-\overline{X_j}\rvert > W_p$
Duncan	$D_p = \begin{cases} Q_{\alpha,p,v}\sqrt{\frac{HKO}{n}}, & n_1 = n_2 = \cdots = n_k = n \\ Q_{\alpha,p,v}\sqrt{\frac{HKO}{k}\left(\frac{1}{n_i}+\frac{1}{n_j}+..+\frac{1}{n_k}\right)}, & n_1 \neq n_2 \neq \cdots \neq n_k \end{cases}$ $v = (n-k)$	$\lvert\overline{X_i}-\overline{X_j}\rvert > D_p$
R-E-G-W-F	F = $\frac{\left(\sum_{i\epsilon R} n_i\overline{x_i}^2 - (\sum_{i\epsilon R} n_i\overline{x_i})^2/\sum_{i\epsilon R} n_i\right)}{(r-1)HKO}$ $r = j - i + 1$, R={$i, \ldots, j$} $\gamma_r = \begin{cases} 1-(1-\alpha)^{r/k} & r < k-1 \\ \alpha & r \geq k-1 \end{cases}$ $f = \sum_{i=1}^{k}(n_i - 1)$	F ≥ $F_{\gamma_r,(r-1),f}$
R-E-G-W-Q	Q= $\begin{cases} \frac{\overline{X_i}-\overline{X_j}}{\sqrt{\frac{HKO}{n}}}, & n_1 = n_2 = \cdots = n_k = n \\ \frac{\overline{X_i}-\overline{X_j}}{\sqrt{\frac{HKO}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}}, & n_1 \neq n_2 \neq \cdots \neq n_k \end{cases}$ $v = (n-k)$	Q > $q_{\alpha,k,v}$
Waller Duncan	$W = t_B$ *(w,f,(k-1),F)* $\sqrt{\frac{2HKO}{n}}$ f = k(n-1) F = $GAKO/HKO$	$\lvert\overline{X_i}-\overline{X_j}\rvert > W$

Burada;

t_B *(w,f,(k-1),F)*: Waller Duncan (Bayesci yaklaşım t) tablo değeri

$q_{\alpha,k,v}$, $q_{\alpha,p,v}$: Tukey ve SNK tablo değerini

$Q_{\alpha,p,v}$:Duncan tablo değerini

$F_{\gamma_r,(r-1),f}$: F tablo değerini göstermektedir.

Çizelge 2.4’te belirtilen karar kriterleri sağlandığı takdirde yokluk hipotezi reddedilerek gruplar arasında farklılık olduğuna karar verilir.

2.2.3. Varyanslar Homojen Olduğunda Kullanılan Çoklu Karşılaştırma Testleri

Varyanslar homojen olduğunda kullanılan çoklu karşılaştırma test grubundan LSD testinin en büyük dezavantajı birinci tip hataya karşı oldukça korunmasız olmasıdır. O yüzden fazla grup sayısı söz konusu olduğunda çok fazla tercih edilmez. Sidak testi özellikle LSD’nin barındırmış olduğu hatayı geliştirmek üzerine kurulmuştur ve hata miktarlarına karşı sınırlamalar getirmektedir (Kayri 2009).

Bonferroni testide Tukey testi gibi çok fazla tercih edilen bir karşılaştırma yöntemidir. Student t istatistiğini temel almaktadır. Çoklu karşılaştırma testlerinde kullanılan önem seviyesini düzenlemektedir. Dunnett testi ise diğerlerinden farklı olarak bir kontrol grubu kullanmaktadır. Testin her aşamasında grup ortalamalarını kontrol grubu ile karşılaştırarak hipotezler hakkında karar vermektedir (SPSS 2011).

Bahsedilen çoklu karşılaştırma testleri Çizelge 2.5’te gösterilmiştir (Carmer and Walker 1985, SPSS 2011).

Çizelge 2.5 Varyanslar Homojen Olduğunda Kullanılan Çoklu Karşılaştırma Testleri

TEST	KRİTİK DEĞER	KARAR
Bonferroni	t= $\frac{\overline{X_i}-\overline{X_j}}{\sqrt{HKO\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}}$, $n_1 = n_2 = \cdots = n_k = n$ $v = (n-k)$	t > $t_{\alpha/k,v}$
Dunnett	DT= $\begin{cases} \frac{\overline{X_i}-\overline{X_c}}{\sqrt{\frac{2HKO}{n}}}, & n_1 = n_2 = \cdots = n_k = n \\ \frac{\overline{X_i}-\overline{X_c}}{\sqrt{\frac{2HKO}{k}\left(\frac{1}{n_i}+\frac{1}{n_j}+\cdots+\frac{1}{n_k}\right)}}, & n_1 \neq n_2 \neq \cdots \neq n_k \end{cases}$ $v = (n-k)$	DT > $Q''_{\alpha,(k-1),v}$
LSD	LSD= $\begin{cases} t_{\alpha,v} \sqrt{\frac{2HKO}{n}}, & n_1 = n_2 = \cdots = n_k = n \\ t_{\alpha,v} \sqrt{HKO\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}, & n_1 \neq n_2 \neq \cdots \neq n_k \end{cases}$ $v = (n-k)$	$\lvert \overline{X}_i - \overline{X}_j \rvert > LSD$
Dunn-Sidak	G.A=$(\overline{X}_i - \overline{X}_j) \pm\ t_{\alpha',v}\sqrt{HKO\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$ $\alpha' = \frac{1}{2}(1-(1-\alpha)^{1/k})$ $v = (n-k)$	G.A ‘0’ değerini içermiyorsa

Burada;

$t_{\alpha/k,v}$, $t_{\alpha,v}$, $t_{\alpha',v}$: Student t tablo değerini

$Q''_{\alpha,(k-1),v}$: Dunnett tablo değerini göstermektedir.

Çizelge 2.5’te belirtilen karar kriterleri sağlandığı takdirde yokluk hipotezi reddedilerek gruplar arasında farklılık olduğuna karar verilir.

Uygulamalarda çoğu zaman üzerinde durulan özellik bakımından elde edilen verilerde varyansların homojenliği ön şartının sağlanamadığı durumlarla sık karşılaşılmaktadır. Bu gibi durumlar söz konusu olduğu zaman genellikle ya veriler uygun bir transformasyona tabii tutulur ya da varyans analizinin parametrik olmayan karşılığı olan Kruskal-Wallis testi kullanılır (Mendeş 2002). Literatürde varyans analizine alternatif uygulanabilecek birçok yöntem bulunmaktadır. Bu çalışmada transformasyona tabii tutulan verilere varyans analizi uygulanması sonucu kullanılan çoklu karşılaştırma testlerinden bahsedilmiştir.

2.2.4. Varyanslar Heterojen Olduğunda Kullanılacak Testler

Varyanslar heterojen olduğunda kullanılan karşılaştırma testi grubundan Dunnett T3 testi student maksimum modulus tablo değerin temel almakta ve Games Howell testine göre daha küçük örneklem büyüklüklerinde kullanılmaktadır. Games Howell testide Dunnet T3 testine benzerdir fakat student range istatistiğini kullanmaktadır ve daha güçlüdür. Dunnett C testi de aynı şekilde karşılaştırmalarda student range istatisiğini kullanmaktadır (SPSS 2011).

Bahsedilen çoklu karşılaştırma testleri Çizelge 2.6'da gösterilmiştir (Üçkardeş 2006, SPSS 2011).

Çizelge 2.6 Varyanslar Heterojen Olduğunda Kullanılan Testler

TEST	KRİTİK DEĞER	KARAR
Dunnett T3	G.A=$(\bar{X}_i - \bar{X}_j) \pm m_{\alpha,k^*,v_{ij}}\sqrt{(S_i^2/n_i + S_j^2/n_j)}$ v_{ij}=$(S_i^2/n_i + S_j^2/n_j)^{2}/(S_i^4/n_i^2V_i + S_j^4/n_j^2V_j)$ $V_i = n_i - 1$, $V_j = n_j - 1$, $k^* = k(k-1)/2$	G.A '0' değerini içermiyor ise
Games -Howell	G.A=$(\bar{X}_i - \bar{X}_j) \pm q_{\alpha,k^*,v_{ij}}\sqrt{(S_i^2/n_i + S_j^2/n_j)\ /2}$ V_{ij}=$(S_i^2/n_i + S_j^2/n_j)^{2}/(S_i^4/n_i^2V_i + S_j^4/n_j^2V_j)$ $V_i = n_i - 1$, $V_j = n_j - 1$, $k^* = k(k-1)/2$	G.A '0' değerini içermiyor ise
Dunnett C	G.A= $(\bar{X}_i - \bar{X}_j) \pm Q/\sqrt{2}$ Q=A/B , B= $\sqrt{(S_i^2/n_i + S_j^2/n_j)}$, A=($q_{\alpha,k,v_i}\ S_i^2/n_i + q_{\alpha,k,v_j}\ S_j^2/n_j)$, $V_i = n_i - 1$, $V_j = n_j - 1$	G.A '0' değerini içermiyor ise

Burada;

$m_{\alpha,k^*,v_{ij}}$: Student Makismum Modulus tablo değerini

$q_{\alpha,k^*,v_{ij}}$: Tukey ve SNK kritik tablo değerini

q_{α,k,v_i} :Tukey ve SNK kritik tablo değerini göstermektedir.

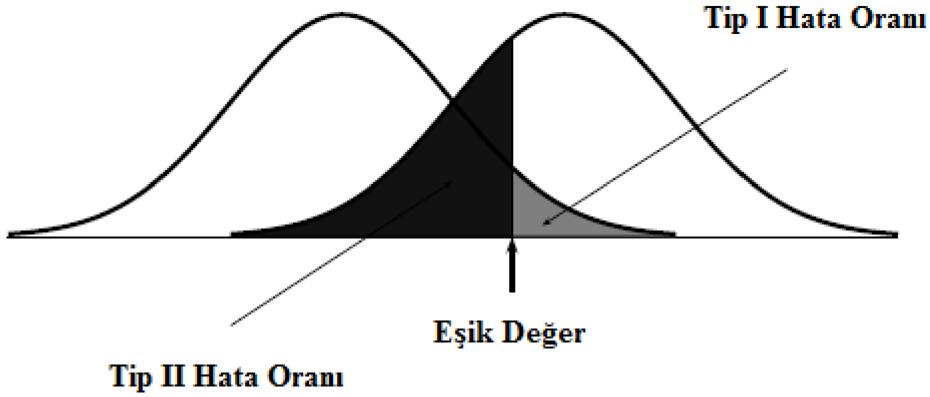
Çizelge 2.6'da belirtilen karar kriterleri sağlandığı takdirde yokluk hipotezi reddedilerek gruplar arasında farklılık olduğuna karar verilir.

2.3. ARDIŞIK ÇOKLU KARŞILAŞTIRMA TESTLERİ

Üç ya da daha fazla grubun yer aldığı denemelerde gruplar arasında farklılığı yaratan grup ya da grupların belirlenmesi ile ilgili çalışmalarda araştırmacılar, dikkate alacakları hata birimini (karşılaştırma başına hata oranı (comparisonwise error rate) ya da deneysel ortak hata oranı (familywise error rate, (FWER)) 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 (n) artması ile paralel olarak değerinin yaklaşık $1-(1-\alpha)^n$ 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 (Cribbie and Keselman 2003). Bundan dolayı deneysel ortak hata oranının kontrol edilmesi yaklaşımı hemen hemen tüm araştırmalarda tercih edilebilecek ortak bir yaklaşımdır (Doğan ve Doğan 2013)

Çoklu hipotez testi problemleri çok karmaşık bir yapıya sahiptir. Her hipotezin testi için Tip I ve Tip II hata söz konusudur fakat tümel hata oranı olarak hangisinin kullanılacağı belirgin değildir. Bu konuda önerilen hata oranı hipotez ailesindeki bir ya da daha fazla hipotezin yanlışlıkla reddedilmesi olasılığı olan FWER'dir. Tercih edilen her test için α anlamlılık seviyesinde Tip I hata oranını kontrol etmek yerine, α anlamlılık seviyesinde tüm hipotez ailesine ait hata oranı FWER'in kontrolüdür. FWER tüm karşılaştırmalar için FWER $\leq \alpha$ koşulunu sağlayarak anlamlılık seviyesini koruduğundan dolayı oldukça güçlüdür (Storey 2002).

Bir hipotez testinde, ilgili alternatifler için kabul edilebilir Tip II hatayı göze aldığımızda, Tip I hatayı kontrol etmek için hipotezin reddedilmesindeki eşik değer değerinin belirlenmesi Şekil 1.'de gösterilmeye çalışılmıştır.



Şekil 2.1 Tip I, Tip II hata oranları ve eşik değerin gösterimine ilişkin grafik (İnt.Kay.1)

Burada eşik değerin seçiminde her bir karşılaştırma için Tip I hatanın, deneysel hata oranı ve yanlış bulgu oranının (FDR) kontrol edilmesi gerekir (İnt. Kay. 1)

Ardışık çoklu karşılaştırma testleri hipotezleri test ederken yokluk hipotezlerine karşılık gelen $p_{(1)}$, $p_{(2)}$, … , $p_{(m)}$ şeklinde sıralanmış p değerlerini dikkate alır ve her bir karşılaştırma için yeniden hesaplanan farklı bir anlamlılık seviyesi kullanır. Ardışık çoklu karşılaştırma testleri uygulanışlarına göre ikiye ayrılır bunlar; azalan aşamalı ve artan aşamalı yöntemlerdir. Artan aşamalı yöntemlerde yokluk hipotezlerine karşılık gelen $p_{(1)}$, $p_{(2)}$, … , $p_{(m)}$ şeklindeki sıralı p değerlerinden en büyük $p_{(m)}$ değerinden başlanarak en küçük değere doğru hipotezler sırasıyla test edilirken azalan aşamalı yöntemlerde sıralanmış değerler içinden en küçük p değeri ile başlanarak en büyüğüne doğru sıra ile yöntem yürütülür (Nichols and Hayasaka 2003).

Bu bölümde ardışık çoklu karşılaştırma testlerinden artan ve azalan aşamalı yöntemler ile bunların temelini oluşturan Bonferroni yönteminden bahsedilmiştir.

2.3.1. Bonferroni Yöntemi

Bonferroni yöntemi deneysel ortak hata onarını koruma altına alan;

$$\mathrm{P}\{\cup_{i=1}^{m}(p_i \leq \alpha/m\} \leq \alpha \qquad (2.6)$$

Bonferroni eşitsizliği üzerine kurulmuştur.

$H_0 = \{H_{(1)}, H_{(2)}, \ldots, H_{(m)}\}$ yokluk hipotezlerine sırasıyla karşılık gelen ve p değerlerini gösteren; $p_{(1)}, p_{(2)}, \ldots, p_{(m)}$ değerleri tanımlı olsun. Tüm hipotezler sırasıyla;

$$p_i \leq \alpha/m \qquad (2.7)$$

Koşulu sağlanırsa reddedilir ve diğer hipoteze geçilir. Diğer durumda kabul edilerek yöntem bitirilir (Hochberg 1988).

2.3.2. Holm Yöntemi

Holm (1979), önerdiği yöntemde Bonferroni yöntemi gibi anlamlılık seviyelerinin her bir adımda farklı hesaplandığını göstermiştir.

$H_0 = \{H_{(1)}, H_{(2)}, \ldots, H_{(m)}\}$ yokluk hipotezlerine sırasıyla karşılık gelen ve p değerlerini gösteren; $p_{(1)}, p_{(2)}, \ldots, p_{(m)}$ değerleri,

$$p_{(1)} \geq p_{(2)} \geq, \ldots, \geq p_{(m)} \qquad (2.8)$$

olacak şekilde en büyük değerden en küçük değere doğru sıralanır. İşleme ilk önce en büyük p değeri ile başlanır.

$$p_{(1)} \leq \alpha/m \qquad (2.9)$$

p değeri hesaplanan anlam düzeyinden küçük ise hipotez reddedilir diğer hipoteze geçilir. Sonraki adımda;

$$p_{(2)} \leq \alpha/(m-1) \qquad (2.10)$$

p değeri hesaplanan anlam düzeyinden küçük ise hipotez reddedilir değilse işlem bitirilir. Bu süreç böyle devam ettirilir. En son aşamada;

$$p_{(m)} \leq \alpha/1 \qquad (2.11)$$

Koşuluna bakılır ve işlem sonlandırılır.

Klasik Bonferroni yönteminin kullanıldığı bütün durumlarda kullanılabilir. Holm yönteminde güç önemli hipotezlere yönlendirilir böylece bir güç artışı sağlanır (Holm 1979, Doğan ve Doğan 2013).

2.3.3. Shaffer Yöntemi

Shaffer (1986), Holm yöntemine katkıda bulunan yeni bir yöntem sunmuştur. Holm yöntemindeki $(i-1)$ hipotezin reddedildiği göz önüne alındığında, paydada yer alan mümkün doğru hipotez sayısının, $(m-i+1)$ daha az hesaplanabileceğini göstermiştir. Ve testin her bir aşamasındaki maksimum doğru yokluk hipotezlerin sayısını veren bir tablo hazırlayıp sunmuştur (Doğan ve Doğan 2013).

Bu yönteme göre;

$H_0 = \{H_{(1)}, H_{(2)}, \ldots, H_{(m)}\}$ yokluk hipotezlerine sırasıyla karşılık gelen ve p değerlerini gösteren; $p_{(1)}, p_{(2)}, \ldots, p_{(m)}$ değerleri,

$$p_{(1)} \leq p_{(2)} \leq, \ldots, \leq p_{(m)} \quad (2.12)$$

Olacak şekilde en küçükten değerden en büyük değere doğru sıralanır.

Hipotezlerden $(i-1)$ tanesinin yanlış olduğunu varsayıldığı durumda maksimum doğru hipotez sayısı t_i şeklinde tanımlanmaktadır. t_i değeri i'nin bütün adımları için ayrı ayrı hesaplanır.

$$p_{(i)} > \alpha / t_i \quad (2.13)$$

p değeri hesaplanan anlamlılık seviyesinden küçük ise hipotez reddedilir ve aynı şekilde yöntem devam ettirilir. Koşul sağlanırsa hipotez kabul edilir ve yöntem bitirilir. $H_{(1)}$, $H_{(2)}, \ldots, H_{(i-1)}$ tane hipotez reddedilmiş olur.

Shaffer yönteminde, maksimum mümkün doğru hipotez sayısı

$$t_i < (m-i+1) \quad (2.14)$$

Olduğundan dolayı deneysel anlamlılık düzeyini korur ve birçok uygulamada daha güçlüdür (Shaffer 1986).

2.3.4.Holland-Copenhaver Yöntemi

Holland-Copenhaver (1987), Holm ve Shaffer yöntemlerinden daha güçlü bir yöntem sunmuştur. Karşılaştırılacak m tane $\{H_{(1)}, H_{(2)}, ... , H_{(m)}\}$ hipotezin testinde kullanılacak

$$C(m) = 1 - (1 - \alpha)^{1/m} \quad (2.15)$$

Eşitsizliğini tanımlamıştır. Yönteme göre;

$H_0 = \{H_{(1)}, H_{(2)}, ... , H_{(m)}\}$ yokluk hipotezlerine sırasıyla karşılık gelen ve p değerlerini gösteren; $p_{(1)}, p_{(2)}, \dots , p_{(m)}$ değerleri,

$$p_{(1)} \leq p_{(2)} \leq , ... , \leq p_{(m)} \quad (2.16)$$

Olacak şekilde en küçükten değerden en büyük değere doğru sıralanır.

$$p_{(i)} > 1 - (1 - \alpha)^{1/(m-i+1)} \quad (2.17)$$

$H_{(1)}, H_{(2)}, ... , H_{(i-1)}$ hipotezleri reddedilir ve geri kalan $H_{(i)}, ... , H_{(m)}$ hipotezleri kabul edilir.

$(m \geq 1)$ iken

$$1 - (1 - \alpha)^{1/m} \geq \alpha/m \quad (2.18)$$

olduğu için Holm ve Shaffer yöntemlerine göre daha fazla hipotezi reddeder (Holland and Copenhaver 1987).

2.3.5.Hochberg Yöntemi

Hochberg (1988), Simes (1986) eşitsizliği üzerine kurulmuş bir yöntemle deneysel ortak hata oranını daha güçlü bir şekilde kontrol altında tuttuğunu göstermiştir. Yönteme göre;

$H_0 = \{H_{(1)}, H_{(2)}, ... , H_{(m)}\}$ yokluk hipotezlerine sırasıyla karşılık gelen ve p değerlerini gösteren; $p_{(1)}, p_{(2)}, \dots , p_{(m)}$ değerleri,

$$p_{(1)} \leq p_{(2)} \leq , ... , \leq p_{(m)} \quad (2.19)$$

Olacak şekilde en küçükten değerden en büyük değere doğru sıralanır. En büyük $p_{(m)}$ değerinden başlanarak hipotezler test edilir.

Bütün $i = m, m - 1, ... , 1$ değerleri için

$$p_{(i)} \leq \alpha/(m - i + 1) \quad (2.20)$$

p değeri hesaplanan anlamlılık seviyesinden küçük ise $H_{(1)}$, $H_{(2)}$, ... , $H_{(i-1)}$ hipotezleri reddedilir ve $H_{(i)}$, ... , $H_{(m)}$ hipotezleri kabul edilir. Tersi durum söz konusu olana kadar yöntem bu şekilde sürdürülerek hipotezler test edilir.

Holm yöntemi ile benzer aşamalarına sahiptir fakat burada en büyük p değerinden başlandığı için Holm yöntemine göre daha fazla hipotezin incelenmesi mümkündür (Hochberg 1988).

2.3.6.Hommel Yöntemi

Hommel(1988), Simes (1986) sunulan metodun üzerine kurulu, hesaplamada testlerin sırasının yanı sıra p değerlerini de göz önünde bulunduran bir yöntem önermiştir. Yöntem iki aşamada gerçekleşir;

$H_0 = \{H_{(1)}, H_{(2)}, \dots, H_{(m)}\}$ yokluk hipotezlerine sırasıyla karşılık gelen ve p değerlerini gösteren; $p_{(1)}, p_{(2)}, \dots, p_{(m)}$ değerleri,

$$p_{(1)} \leq p_{(2)} \leq, \dots, \leq p_{(m)} \tag{2.21}$$

Olacak şekilde en küçükten değerden en büyük değere doğru sıralanır.

$$J = \max\{i \epsilon \{1,2,\dots,m\}: p_{(m-i+k)} > k\alpha/i; k = 1,2,\dots,i\} \tag{2.22}$$

İlk aşamada $i = 1, k = 1$ alınarak işleme başlanır. Birinci adım yalnızca bir test içerir.

$$p_{(m-i+k)} > k\alpha/i \tag{2.23}$$

Koşul sağlanırsa işlem devam ettirilir.

İkinci adımda $i = 2, k = 1$ alında ve $i = 2, k = 2$ alındığında p değerleri kontrol edilir. (2.18) koşulu sağlarsa işlem devam ettirilir.

Süreç bu şekilde $i = 1,2,\dots,m$ için devam eder. (2.18) koşulu sağlanmadığı durumda birinci aşamaya son verilir. En büyük J değeri belirlenmiş olur.

$$maxJ = j' \tag{2.24}$$

İkinci aşamada düzeltilmiş anlamlılık seviyesi belirlenmiş olarak $p_{(1)} \leq p_{(2)} \leq, \dots, \leq p_{(m)}$ en büyük p değerine sahip hipotezden başlanarak bütün hipotezler test edilir.

$$p_{(i)} \leq \alpha/j' \tag{2.25}$$

$p_{(i)}$ değeri düzeltilmiş anlamlılık seviyesinden küçük olduğu sürece bütün $H_{(i)}$ hipotezleri reddedilir. Diğer durumda işlem sonlandırılır. (Hommel 1988, Doğan ve Doğan 2013).

2.3.7.Rom Yöntemi

Rom(1990), Hochberg yöntemine benzer fakat hipotezlerin farklı bir anlamlılık seviyelerinde daha güçlü bir şekilde test edildiği bir yöntem önermiştir.

Yönteme göre;

$H_0 = \{H_{(1)}, H_{(2)}, \dots , H_{(m)}\}$ yokluk hipotezlerine sırasıyla karşılık gelen ve p değerlerini gösteren; $p_{(1)}, p_{(2)}, \dots , p_{(m)}$ değerleri,

$$p_{(1)} \leq p_{(2)} \leq , \dots , \leq p_{(m)} \qquad (2.26)$$

Olacak şekilde en küçükten değerden en büyük değere doğru sıralanır. Hochberg yönteminden farklı olarak yeni bir anlamlılık seviyesi hesaplanır.

Bütün $i = m, m-1, \dots, 1$ değerleri için;

$$\alpha'_{(m-i+1)} = \left[\sum_{j=1}^{i-1} \alpha^j - \sum_{j=1}^{i-2} \binom{i}{j} \alpha'^{(i-j)}_{(m-j)}\right] / i \qquad (2.27)$$

Şeklinde hesaplanan anlamlılık seviyesi en büyük $p_{(m)}$ değerinden başlanarak her aşamada tekrar hesaplanmak koşulu ile tüm hipotezler test edilir.

$$p_{(i)} \leq \alpha'_{(m-i+1)} \qquad (2.28)$$

$p_{(i)}$ değerleri hesaplanan anlamlılık seviyesinden küçük olduğu sürece $H_{(i)}$ hipotezleri reddedilir, diğer durumda hipotez kabul edilerek işlem bitirilir (Rom 1990, Doğan ve Doğan 2013).

2. 4. FDR Testi

Bonferroni ya da Bonferroni-Holm gibi geleneksel çoklu düzeltmeler, en az bir yanlış kabul etme kararına ait olasılığı olan FWER' i kontrol ederler. Dudoit vd. (2000) çalışmalarında, Gen tanımlama deneylerinde ortak hatadan kaynaklanan bağımlı yapıyı düzelten metotları içeren prosedürlerlerin kontrolünde FWER' i ele almışlardır. FWER kontrol prosedürleri, beklenen en az bir yanlış kabul kararı verilme olasılığı arttırılmadıkça genellikle kullanışlı sonuçlar vermektedir. Biyolojik prosedürleri daha iyi sonuçlandırmak için binlerce gen arasından istenilen bir gen araştırıldığında birçok doğru karara sebep olan bazı yanlış kararlara göz yumulabilir, ancak bu yanlış karar içerisinde bizim araştırdığımız bir genin olması da söz konusu olabilir. Reddedilen tüm kararlar arasında yanlışlıkla reddedilenlerin oranı FDR olarak adlandırılır. Bu oranı kontrol etmede Bonferroni metoduna benzeyen bu metot ilk olarak Benjamini ve Hochberg (1995) tarafından tanıtılmıştır. Storey (2002) çalışmasında FDR'nin kontrolünün, anlamlı (olumlu) kararlar söz konusu olduğunda dikkat çektiğini ifade etmiştir ve PFDR ismi ile mevcut FDR için bu anlamlı kararların dikkate alındığı bir teknik geliştirmiştir. Storey ve Tibshirani (2001) çalışmalarında PFDR'yi tahmin etmede metotlar önermiş ve ortak hata için birçok simülasyon çalışması yapmışlardır. Bunların dışında FWER ya da FDR'yi kontrol etmek için yapılan bazı çalışmalar da söz konusudur. Keselman vd. (2002), Reiner vd. (2003) yaptıkları karşılaştırmalı çalışmalarda FDR'nin FWER'den daha güçlü sonuçlar sağladığını göstermişlerdir (Scheid and Spang, 2003, Doğan ve Doğan 2013).

FWER değerinin kontrol altında tutulması yaklaşımı, araştırmacılar tarafından sıklıkla başvurulan bir yaklaşım olmasına rağmen uygulamalı araştırmalarda karşılaşılan bazı zorlukları bulunmaktadır. Bunlar;

- FWER değerinin kontrol edilmesi ile ilgili metodolojide kullanılan testler çoğunlukla çok değişkenli normal dağılım üzerine kurulu olmasına rağmen gerçekte test istatistikleri çok değişkenli normal değildir.

- Birinci tür hata ile ilgili klasik değerler dikkate alınarak tek tek karşılaştırma yapılması durumunda, FWER değerinin kontrol edilmesinde kullanılan klasik işlemlerin gücü, diğer işlemlerin gücüne göre daha düşüktür.

- FWER değerinin kontrol edilmesine her zaman gerçekten ihtiyaç olmayabilir. Çünkü FWER değerinin kontrol edilmesi, karşılaştırılacak gruplardan en az bir tanesi ile ilgili yanlış olabilecek yorumlar içeren sonuçlar söz konusu olduğunda önemlidir (Benjamini and Hochberg 1995).

Çoklu hipotez testi problemlerinde Tip I hata için tek bir ölçü yoktur. Standart ölçü, herhangi Tip I hata olasılığı olan FWER'dir. Son yıllardaki yeni gelişme FDR hata metriğidir ki bu yanlış pozitiflerin (kabullerin) arasından red edilen hipotezlerin beklenen oranıdır. FDR prosedürleri, yanlış kabullerin kontrolünü kullanışlı bir açıdan kontrol ettiğinden FEW'den daha güçlüdür (Nichols and Hayasaka 2003).

FDR ile ilgili daha geniş bilgi çalışmanın Meteryal ve Metot kısmında sunulmuştur.

3. MATERYAL ve METOT

Çalışmanın uygulama kısmında çoklu karşılaştırma testlerine alternatif bir teknik olan FDR testinin performansı incelenmiştir. Çalışmada, 3, 5 ve 10 grubun söz konusu olması durumunda 50, 100 ve 200 birimlik örneklemler için FDR sonucundan elde edilen anlamlılık değeri ile t testinden elde edilen anlamlılık değerleri karşılaştırılmıştır. İlgili verilerin türetilmesi ve çözümlemelerin gerçekleştirilmesi için MATLAB programından yararlanılmıştır. MATLAB programında yapılan veri türetimi, çözümleme ve simülasyonlar için yazılan kodlar Ek 1, Ek 2 ve Ek 3'te verilmiştir. Eklerde verilen kodlar 3 grup söz konusu olduğu durumlar için geçerli olup aynı prosedür 5 ve 10 grup için de uygulanmıştır. Tekniklere ait kodların içerisinde MATLAB'ın kendi sitesinde yer alan FDR analizi için yer alan kodlardan da yararlanılmıştır.

Çoklu önemlilik testleri ile ilgili bazı zorlukları gidermek üzere farklı bir yaklaşım önerilmektedir. Yanlış bulgu oranı (False Discovery Rate, FDR) olarak isimlendirilen bu yaklaşım, yanlışlıkla reddedilen hipotezlerin beklenen oranı olarak ifade edilmektedir. FDR tüm hipotezler doğru olduğunda FWER değerine eşit olmaktadır. Üstelik hipotezlerden en az bir tanesinin doğru olmaması durumunda FDR değeri FWER değerinden daha küçük olmakta, dolayısıyla da istatistiksel gücü artırdığından FWER yerine FDR' nin kullanılması daha çok arzu edilmektedir. FDR ve FDR'nin son zamanlardaki güncellenmiş hallerinin istatistiki açıdan anlamı, klasik çoklu karşılaştırmalarda p değerinin birleştirilmesi temeline dayanır (Yudi *et al*. 2005).
FDR testi geleneksel FWER kontrol metoduna göre daha güçlü çoklu hipotez testi kriteri sağlamaktadır. (Benjamini and Hochberg, 1995). Bundan dolayı FDR metodu Genome-Wide Association (GWA) (büyük gen kuruluşu) tarafından yoğun bir biçimde binlerce SNP'nin test edilmesinde kullanılmıştır. FDR kontrolü, tüm hipotezler için p değerlerinin toplanımı ile gerçekleştirilmektedir ve p değerleri için kritik değerler Doğru olan alternatif hipotez dağılımına bağlı olarak sabit FDR kontrol seviyesine göre değişir. FDR Metodu ayrıca FDR ye göre düzeltilmiş p değerlerini, örneğin her bir hipotez testi için q değerlerini elde etmek ve bu q değerlerini doğrudan FDR kontrol seviyesi ile test etmek suretiyle de uygulanabilir (Storey 2002, Doğan ve Doğan 2013).

Bu manada “Discovery” kelimesi ilk kez Soriç (1989) tarafından ortaya atılmış ve bir yokluk hipotezinin geçici olarak reddi ya da bir alternatif hipotezin geçici olarak kabulü olarak ifade edilmiştir. Soriç (1989)’e göre tek yanlı bir testte yokluk hipotezinin geçici olarak reddi ya da tek yanlı bir güven aralığı sıfırdan farklı bir etki göstermektedir. r tane yokluk hipotezinin yanlışlıkla reddedildiği m tane bağımsız deneme için hata oranının üst sınırı;

$$Q_{\max} = \left(\frac{m}{r} - 1\right)\alpha/(1-\alpha) < 1, \tag{3.1}$$

r tane yanlış bulguya ait güven aralığı için hata oranı (FDR) ise;

$$E = \alpha \mathrm{m}/\mathrm{r} \tag{3.2}$$

Şeklinde reddedilen tüm kararlar arasında yanlışlıkla reddedilenlerin beklenen oranını ifade eder. Çizelge 3.1’den de görüleceği üzere;

V: Yanlış bulguların sayısını ifade eden rassal bir değişken,

R: Çoklu test prosedüründen elde edilen anlamlı sonuç sayısını ifade etmek üzere, Benjamini ve Hochberg FDR’yi şöyle tanımlamıştır.

R>0 olduğunda;

$$\mathrm{FDR} = E(V/R) \tag{3.3}$$

Şeklinde ifade edilir. Diğer durumlarda 0 ‘dır.

Microarray gibi büyük ölçekteki hipotez testleri için FDR, en az bir yanlış bulgunun olasılığı olarak tanımlanan FWER’den daha uygun görünmektedir (Hochberg and Tamhane 1987, Cyril *et al.*2005).

R değeri, reddedilen toplam yokluk hipotezi sayısını göstermek üzere m_0 tanesi doğru olan m tane yokluk hipotezinin eşanlı olarak test edildiği durumda ortaya çıkabilecek muhtemel sonuçlar özet olarak Çizelge 3.1’de verilmiştir (Doğan ve Doğan 2013).

Çizelge 3.1 m Tane Eşanlı Yokluk Hipotezi Testinden Elde Edilebilecek Hata Sayıları

		Yokluk Hipotezi Önemsiz (Kabul)	Önemli (Ret)	Genel
Yokluk	Doğru	U	V	m_0
Hipotezi	Yanlış	T	S	m_1
	Genel	m-R	R	M

Çizelge 3.1'de yer alan R gözlemlenebilen rasgele değişken, U, V, S ve T ise gözlemlenemeyen rasgele değişkenlerdir. Her bir yokluk hipotezi ayrı ayrı α anlamlılık düzeyinde test edilirse R=R(α) değeri artar. Çizelge 3.1.'de yer alan değerlerden yararlanarak;

Karşılaştırma başına hata oranı= E(V/m) (3.4)

Deneysel ortak hata oranı= $P(V \geq 1)$ (3.5)

olarak ifade edilir.

Her bir yokluk hipotezinin ayrı ayrı α anlamlılık düzeyinde test edilmesi

$E(V/m) \leq \alpha$, (3.6)

Her bir yokluk hipotezinin α/m anlamlılık düzeyinde test edilmesi ise

$P(V \geq 1) \leq \alpha$ (3.7)

olmasını garanti eder (Benjamini and Hochberg, 1995).

Reddedilen yokluk hipotezleri içerisinde yanlışlıkla reddedilen yokluk hipotezlerinden kaynaklanan hata oranı;

$Q = V/(V+S)$ (3.8)

biçiminde ifade edilir.

Doğal olarak, V+S sıfır ise Q = 0'dır.

Q gözlemlenemeyen rasgele değişkendir. Çünkü V ve S bilinmemektedir. FDR değeri Q'nun bekleneni olmaktadır ve Q_e ile ifade edilmektedir. Dolayısıyla;

$FDR = Q_e = E(Q) = E\{V/(V+S)\} = E(V/R)$'dir. (3.9)

FDR ile ilgili iki önemli özellik söz konusudur. (Benjamini and Hochberg, 1995, Doğan ve Doğan 2013). Bunlar;

- $m_0 = m$ olduğunda yani yokluk hipotezlerinin tamamı doğru ise S = 0 ve V=R durumunda FDR değeri FWER değerine eşittir.
  Eğer V = 0 ise o zaman Q = 0 olur. Eğer V > 0 ise o zaman Q = 1 olur.
  Bu durumda; $P(V \geq 1) = E(Q) = Q_e$
  Bundan dolayı FDR'nin kontrolü zayıf da olsa FWER'in kontrolü demektir.
- $m_0 < m$ olduğunda yani en az bir yalnış hipotez varsa, FDR değeri FWER değerine ya eşit ya da küçüktür aynı zamanda FDR ile FWER değerleri birbirlerinden oldukça farklı olabilir.
  Eğer V > 0 ise $V/R \leq 1$ bu durumda da $P(V \geq 1) \geq Q_e$ olur.

Sonuç olarak FWER değerini kontrol eden herhangi bir işlem aynı zamanda FDR değerini de kontrol etmektedir. Bununla birlikte, eğer bir işlem yalnızca FDR değerini kontrol ediyorsa FWER değeri de kontrol ediliyor denemez ancak bu durumda güç değerinin artması beklenebilir (Benjamini and Hochberg, 2000).

$H_0 = \{H_{(1)}, H_{(2)}, \ldots\ldots\ldots., H_{(m)}\}$ yokluk hipotezlerine sırasıyla karşılık gelen ve *p* değerlerini gösteren; $p_{(1)}, p_{(2)}, \ldots\ldots\ldots\ldots, p_{(m)}$ değerleri,
$p_{(1)} \leq p_{(2)} \leq, \ldots\ldots\ldots, \leq p_{(m)}$ olacak şekilde sıralansın.

$$P_i \leq \frac{i}{m} q^* \qquad (3.10)$$

eşitsizliğini sağlayan en büyük *i* değeri *k* olsun. Bu durumda tüm $H_{(i)}$ i= 1, 2, …, k hipotezleri reddedilir.
Bu işlem $q^* = \alpha$ yanılma düzeyinde FDR değerini kontrol eden işlem olarak ifade edilir (Benjamini and Hochberg, 1995, Benjamini and Hochberg, 2000).

Eşitsizlik (3.10)'dan yararlanarak;
Herhangi bir $0 \leq m_0 \leq m$ tane doğru yokluk hipotezlerine karşılık gelen bağımsız *p* değerleri ile,
$m_1 = m - m_0$ tane yanlış yokluk hipotezine karşılık gelen *p* değerlerinden yararlanarak;

$$E\left(Q \middle| P_{m_0+1} = p_1, \ldots\ldots\ldots\ldots\ldots\ldots., P_m = P_{m1}\right) \leq \frac{m_0}{m} q^* \qquad (3.11)$$

yazılabilir.
$m_1 = m - m_0$ tane yanlış yokluk hipotezi olduğu düşünüldüğünde Eşitsizlik (3.11)'den yararlanarak;

$$E(Q) \leq \frac{m_0}{m} q^* \leq q^* \quad (3.12)$$

yazılır ve böylece FDR kontrol edilir (Benjamini and Hochberg, 1995 Doğan ve Doğan 2013).

FDR ile ilgili süreç maddeler halinde aşağıdaki gibidir;

$H_0 = \{H_{(1)}, H_{(2)}, \ldots\ldots\ldots., H_{(m)}\}$ yokluk hipotezlerine sırasıyla karşılık gelen ve *p* değerlerini gösteren $p_{(1)}, p_{(2)}, \ldots\ldots\ldots\ldots, p_{(m)}$ değerleri, $p_{(1)} \leq p_{(2)} \leq, \ldots\ldots\ldots., \leq p_{(m)}$ olacak şekilde sıralanır,

$$\hat{k} = \max\{1 \leq k \leq m : \ p_k \leq q^* k/m\} \quad (3.13)$$

değeri belirlenir.

Eğer $\hat{k}$ değeri varsa $p_{(1)} \leq p_{(2)} \leq, \ldots\ldots\ldots., \leq p_{(k)}$ yokluk hipotezleri reddedilir, tersi durumda tüm yokluk hipotezleri kabul edilir (Storey *et al.*, 2004, Doğan ve Doğan 2013).

Bu bölümde FDR testinin özelliklerinin ardından FDR testi, çoklu ve ardışık çoklu karşılaştırma testleri p değerlerini hepsinin bir arada karşılaştırılabileceği bir örnek sunulmuştur.

Örnek: Bir işletmede bulunan 4 farklı makinenin ortalama üretim miktarları arasında farklılık olup olmadığı α= 0,05 önem seviyesinde karşılaştırılmıştır.

Çizelge 3.2 Makinelerin Ortalama Üretim Miktarı Gözlem Değerleri

Makine	Gözlem Değerleri									
A	22	16	16	19	21	18	17	23	20	21
B	30	15	28	18	23	17	20	19	28	25
C	31	26	25	29	30	21	28	22	32	19
D	16	21	20	15	12	21	20	11	18	16

Grup ortalamaları $\bar{X}_A$= 19,3 , $\bar{X}_B = 22{,}3$, $\bar{X}_C = 26{,}3$, $\bar{X}_D = 17$ şeklindedir.

Öncelikle işletmede bulunan makinelerin üretim miktarları arasında fark olup olmadığını belirleyebilmek için Varyans Analizi uygulanmıştır.

H_0: $\mu_A = \mu_B = \mu_C = \mu_D$ (Makinelerin üretim miktarları arasında farklılık yoktur.)

H_1: $\mu_A \neq \mu_B \neq \mu_C \neq \mu_D$ (Makinelerin üretim miktarları arasında farklılık vardır.)

Çizelge 3.3 Makinelerin Ortalama Üretim Miktarı ANOVA tablosu

Değişim kaynağı	(KT)	(SD)	(KO)	F
Gruplar arası	650	3	216,667	
Gruplar içi	606,4	36	16,844	12,863
Genel	1256,4	39		

$$\frac{GAKO}{HKO} = 12{,}863 \geq F_{3,36,0,05} \quad (3.14)$$

Olduğu için H_0 reddedilir ve makinelerin üretim miktarları arasında farklılık olduğuna karar verilir. Hangi makineler arasında farklılık olduğunu belirleyebilmek için çoklu karşılaştırma, ardışık çoklu karşılaştırma testi ve FDR testileri uygulanmıştır.

Bu dört makine için t-testi ile yapılan ikili karşılaştırma sonuçlarından elde edilen p değerleri ardışık ve çoklu karşılaştırma ve FDR testi değerleri ile karşılaştırılmıştır. Aynı veriler kullanılarak dört farklı yöntemden elde edilen sıralanmış p değerlerine ait sonuçlar Çizelge 3.4 de verilmiştir.

Çizelge 3.4 Karşılaştırılan Grupların p, Tukey, Holm ve FDR Değerleri

Karşılaştırılan Gruplar	**t testi p_i değeri**	**Tukey Testi p_i değeri**	**Holm Testi p_i değeri**	**FDR Testi p_i değeri**
$X_C - X_D$	0,000*	0,000*	0,0083*	0,0083*
$X_A - X_C$	0,000*	0,003*	0,01*	0,0166*
$X_B - X_D$	0,017*	0,031*	0,0125	0,025*
$X_B - X_C$	0,083	0,146	0,0166	0,033
$X_A - X_D$	0,116	0,595	0,025	0,0416
$X_A - X_B$	0,120	0,370	0,05	0,05

* Gruplar arasında istatistiksel olarak farklılık olduğunu göstermektedir.

Çizelge 3.4’de görüldüğü gibi Tukey ve t testi sonucunda elde edilen p değerleri α anlamlılık seviyesi ile karşılaştırıldığında ilk üç karşılaştırma yani C-D, A-C ve B-D makineleri arasında farklılık olduğuna karar verilmiştir. Holm testi sonucunda elde edilen değerler ikili karşılaştırma p değerleri ile karşılaştırıldığında C-D ve A-C makineleri arasında farklılık olduğuna karar verilmiştir. FDR testi sonucunda ise elde edilen değerler ikili karşılaştırma p değerleri ile karşılaştırıldığında C-D, A-C ve B-D makineleri arasında farklılık olduğuna karar verilmiştir.

Örnekte görüldüğü gibi küçükten büyüğe doğru sıralanmış p değerlerinden Tukey ve t testlerine ait değerlerin sürekli arttığı gözlenmektedir. Fakat Holm ve FDR testlerine ait p değerleri en fazla anlamlılık seviyesine kadar yükselmektedir. Bu durum ardışık çoklu karşılaştırma ve FDR testlerinde birinci tip hata oranını korunduğunu göstermektedir.

4. BULGULAR

Bu çalışmada, 3, 5 ve 10 grubun söz konusu olması durumunda 50, 100 ve 200 birimlik örneklemler için FDR sonucundan elde edilen anlamlılık değeri ile t testinden elde edilen anlamlılık değerleri karşılaştırılmıştır.

Burada ilk olarak A, B ve C grupları söz konusu olduğunda, gruplar arasında farkın çıkması, ikili karşılaştırma yapmayı gerektireceğinden, MATLAB programında farklı ortalama, farklı standart sapmaya ve farklı örneklem hacimlerine sahip normal dağılıma sahip veri seti oluşturularak çözümlemeler gerçekleştirilmiştir. Grup sayısının üç olduğu durum ele alınmıştır ve üç tane ikili karşılaştırma yapılmıştır. İlgili veri setleri için yapılan analiz sonuçları 50 birimlik örneklem hacmi için Çizelge 4.1.'de, 100 birimlik örneklem hacmi için Çizelge 4.2.'de, 200 birimlik örneklem hacmi için Çizelge 4.3.'te sunulmaktadır.

Çizelge 4.1. Üç grup ve n=50 birimlik örneklem için elde edilen p ve FDR değerleri

Karşılaştırılan Gruplar	**t testi p_i değeri**	**FDR testi p_i değeri**
A-B	0,024055*	0,016667
A-C	0,051984	0,033333
B-C	0,463199	0,050000

* Gruplar arasında istatistiksel olarak farklılık olduğunu göstermektedir.

Örneklem hacminin 50 birim olduğu durumda 3 grup için yapılan 3 ikili karşılaştırma sonuçlarının yer aldığı Çizelge 4.1 incelendiğinde, hem t hem de FDR testi için yapılan ikili karşılaştırmalardan t testine göre ikisinin (A-C, B-C) kabul edildiği, FDR testine göre ise üçünün de kabul edildiği görülmektedir. A-B grubu ikili karşılaştırması için t testi ile 0,02'lik bir oranla reddederek gruplar arasında fark vardır şeklinde karar verilirken, FDR testi ile gruplar arasında farklılık olmadığında karar verilmektedir.

Çizelge 4.2. Üç grup ve n=100 birimlik örneklem için elde edilen p ve FDR değerleri

Karşılaştırılan Gruplar	t testi p_i değeri	FDR testi p_i değeri
A-B	0,002569*	0,016667*
A-C	0,010028*	0,033333*
B-C	0,364527	0,050000

* Gruplar arasında istatistiksel olarak farklılık olduğunu göstermektedir.

Örneklem hacminin 100 birim olduğu durumda 3 grup için yapılan 3 ikili karşılaştırma sonuçlarının yer aldığı Çizelge 4.2 incelendiğinde, hem t hem de FDR testi için yapılan ikili karşılaştırmalardan birinin (B-C) kabul edildiği, diğer iki tanesinin (A-B, A-C) red edildiği görülmektedir. Örneklem hacmi arttıkça t testine ait p değerleri hipotezleri daha küçük olasılıklarla ile reddederken FDR testinde reddetme olasılıkları değişmemiştir.

Çizelge 4.3. Üç grup ve n=200 birimlik örneklem için elde edilen p ve FDR değerleri

Karşılaştırılan Gruplar	t testi p_i değeri	FDR testi p_i değeri
A-B	0,0000114206098*	0,016667*
A-C	0,0002506090000*	0,033333*
B-C	0,2641484300000	0,050000

* Gruplar arasında istatistiksel olarak farklılık olduğunu göstermektedir.

Örneklem hacminin 200 birim olduğu durumda 3 grup için yapılan 3 ikili karşılaştırma sonuçlarının yer aldığı Çizelge 4.3 incelendiğinde, yapılan ikili karşılaştırmalardan ilk ikisinin (A-B, A-C) küçük p değerleri ile red edildiği, diğerinin (B-C) ise kabul edildiği görülmektedir. Burada t testinde p değerleri 100 ve 50 birimlik örneklemler için yapılan karşılaştırmalardaki p değerlerinden çok daha küçük değerler alarak hipotezler reddedilmiştir. A-B grupları arasında t testi ile 0,00001 olasılıkla fark olduğu söylenirken FDR testi ile 0.01 olasılıkla farklılık olduğu ifade edilmektedir. Fakat B-C gruplarına bakıldığında t testi ile 0,26 olasılıkla hipotezler kabul edilirken FDR testi ile 0,05 olasılıkla gruplar arasında farklılık olmadığı ifade edilmiştir. FDR değerlerinin örneklem hacmi artsa bile aynı kaldığı ve 0,05 birinci tip hata oranını aşamadığı görülmektedir.

Grup sayısının 5 olması durumunda ikili karşılaştırma sayısı 10 olmaktadır. Gruplar A, B, C, D ve E olarak düşünüldüğünde ilgili karşılaştırmaların yapılabilmesi için MATLAB programında üç grup söz konusu olduğu durumda olduğu gibi veri setleri oluşturulmuştur. Bu özellikteki veri seti için yapılan analiz sonuçları 50 birimlik örneklem hacmi için Çizelge 4.4.'te, 100 birimlik örneklem hacmi için Çizelge 4.5.'de, 200 birimlik örneklem hacmi için Çizelge 4.6.'da sunulmaktadır.

Çizelge 4.4. Beş grup ve n=50 birimlik örneklem için elde edilen p ve FDR değerleri

Karşılaştırılan Gruplar	**t testi p_i değeri**	**FDR testi p_i değeri**
A-B	0,000034079*	0,005*
A-C	0,000309056*	0,010*
A-D	0,002947086*	0,015*
A-E	0,009709377*	0,020*
B-C	0,026519613*	0,025
B-D	0,053547710	0,030
B-E	0,129254107	0,035
C-D	0,232721538	0,040
C-E	0,433698404	0,045
D-E	0,714021009	0,050

* Gruplar arasında istatistiksel olarak farklılık olduğunu göstermektedir.

Örneklem hacminin 50 birim olduğu durumda 5 grup için yapılan 10 ikili karşılaştırma sonuçlarının yer aldığı Çizelge 4.4 incelendiğinde, yapılan ikili karşılaştırmalardan t testine göre ilk beşinin, FDR testine göre ilk dördünün red edildiği ve diğerlerinin kabul edildiği görülmektedir. A-B ikili karşılaştırması ele alındığında t testinin 0,00003 olasılıkla FDR testinin ise 0,005 olasılıkla hipotezleri reddettiği için t testinin daha hassas olduğu düşünülebilir. Fakat sıraya dizilmiş p değerleri arttıkça FDR testinin güvenilirliği ortaya çıkmaktadır ve anlamlılık seviyesini koruduğu görülmektedir.

Çizelge 4.5. Beş grup ve n=100 birimlik örneklem için elde edilen p ve FDR değerleri

Karşılaştırılan Gruplar	t testi p_i değeri	FDR testi p_i değeri
A-B	0,0000000033*	0,005*
A-C	0,0000007187*	0,010*
A-D	0,0001276159*	0,015*
A-E	0,0006027161*	0,020*
B-C	0,0023604704*	0,025*
B-D	0,0086800038*	0,030*
B-E	0,0311203124*	0,035*
C-D	0,0868255801	0,040
C-E	0,2639783271	0,045
D-E	0,5737905454	0,050

* Gruplar arasında istatistiksel olarak farklılık olduğunu göstermektedir.

Örneklem hacminin 100 olduğu durum için 5 grup için yapılan 10 ikili karşılaştırma sonuçlarının yer aldığı Çizelge 4.5 incelendiğinde, yapılan ikili karşılaştırmalardan t testi ile üçünün kabul edildiği, diğerlerinin 50 birimlik örneklem hacmindeki değerlere göre daha küçük p değerleri ile red edildiği görülmektedir. t testi p değerlerinin ilk yedi karşılaştırmadan sonra anlam seviyesini aştığı gözlemlenmektedir. Aynı şekilde FDR değerlerine bakıldığında ilk yedi tanesinin red edildiği diğerlerinin ise kabul edildiği görülmektedir. FDR değerlerinin örneklem hacmi artsa bile aynı kaldığı ve 0,05 birinci tip hata oranını aşamadığı görülmektedir.

Çizelge 4.6. Beş grup ve n=200 birimlik örneklem için elde edilen p ve FDR değerleri

Karşılaştırılan Gruplar	**t testi p_i değeri**	**FDR testi p_i değeri**
A-B	0,000000000000000022*	0,005*
A-C	0,000000000001040140*	0,010*
A-D	0,000000096212694750*	0,015*
A-E	0,000002088307544477*	0,020*
B-C	0,000027914260987232*	0,025*
B-D	0,000306740919718883*	0,030*
B-E	0,002911819373859250*	0,035*
C-D	0,017417007841139800*	0,040*
C-E	0,089546481681295500	0,045
D-E	0,317129146898809000	0,050

* Gruplar arasında istatistiksel olarak farklılık olduğunu göstermektedir.

Örneklem hacminin 200 birim olduğu durumda 5 grup için yapılan 10 ikili karşılaştırma sonuçlarının yer aldığı Çizelge 4.6 incelendiğinde, yapılan ikili karşılaştırmalardan t testine ikisinin kabul edildiği, diğerlerinin 50 ve 100 birimlik örneklem hacimlerindeki p değerlerine göre küçük değerlerle red edildiği görülmektedir. FDR değerlerine bakıldığında ise aynı şeklide ilk sekiz tanesinin red diğerlerinin kabul edildiği görülmektedir. İlgili p değerlerinin 100 ve 50 birimlik örneklemler için yapılan karşılaştırmalardaki p değerlerinden çok daha küçük olduğu fakat FDR değerlerinin değişmediği göze çarpmaktadır.

Grup sayısının 10 olması durumunda ikili karşılaştırma sayısı 45 olmaktadır. Gruplar A, B, C, D, E, F, G, H, I ve J olarak düşünülerek oluşturulan veri seti için yapılan analiz sonuçları 50 birimlik örneklem hacmi için Çizelge 4.7'de, 100 birimlik örneklem hacmi için Çizelge 4.8'de, 200 birimlik örneklem hacmi için Çizelge 4.9'da sunulmaktadır.

Çizelge 4.7. 10 grup ve n=50 birimlik örneklem için elde edilen p ve FDR değerleri.

Karşılaştırılan Gruplar	**t testi p_i değeri**	**FDR testi p_i değeri**
A-B	0,000000000000000000*	0,001111*
A-C	0,000000000000001240*	0,002222*
A-D	0,000000000000376480*	0,003333*
A-E	0,000000000044204308*	0,004444*
A-F	0,000000000344664272*	0,005556*
A-G	0,000000001815543751*	0,006667*
A-H	0,000000004833152341*	0,007778*
A-I	0,000000016825851582*	0,008889*
A-J	0,000000040869917662*	0,010000*
B-C	0,000000103043573415*	0,011111*
B-D	0,000000270348236957*	0,012222*
B-E	0,000000778189080659*	0,013333*
B-F	0,000001762146871507*	0,014444*
B-G	0,000003708703263271*	0,015556*
B-H	0,000008013586227067*	0,016667*
B-I	0,000017752445402116*	0,017778*
B-J	0,000036269669303618*	0,018889*
C-D	0,000075086434022479*	0,020000*
C-E	0,000140593684139160*	0,021111*
C-F	0,000262173352136893*	0,022222*
C-G	0,000444552272420111*	0,023333*
C-H	0,000793525133591124*	0,024444*
C-I	0,001363556428131020*	0,025556*
C-J	0,002245103594305900*	0,026667*
D-E	0,003633551149332790*	0,027778*
D-F	0,005842341723940550*	0,028889*
D-G	0,009171276566376540*	0,030000*
D-H	0,013966350011112400*	0,031111*
D-I	0,020133176143261600*	0,032222*
D-J	0,028964275642826400*	0,033333*
E-F	0,040024797833452000*	0,034444
E-G	0,054198045658671900	0,035556
E-H	0,072537047198545100	0,036667
E-I	0,094539555660992700	0,037778
E-J	0,121978738596565000	0,038889
F-G	0,156239821648749000	0,040000
F-H	0,195166446608266000	0,041111
F-I	0,244947756432680000	0,042222
F-J	0,305193495512790000	0,043333
G-H	0,370418992588437000	0,044444
G-I	0,450649244469801000	0,045556
G-J	0,542946009703164000	0,046667
H-I	0,647736264875323000	0,047778
H-J	0,759587797346495000	0,048889
I-J	0,883219859196012000	0,050000

* Gruplar arasında istatistiksel olarak farklılık olduğunu göstermektedir.

Çizelge 4.8. 10 grup ve n=100 birimlik örneklem için elde edilen p ve FDR değerleri

Karşılaştırılan Gruplar	**t testi p_i değeri**	**FDR testi p_i değeri**
A-B	0,0000000000000000000000000*	0,001111*
A-C	0,0000000000000000000000000*	0,002222*
A-D	0,0000000000000000000000000*	0,003333*
A-E	0,0000000000000000000000000*	0,004444*
A-F	0,0000000000000000009992007*	0,005556*
A-G	0,0000000000000000193178806*	0,006667*
A-H	0,0000000000000003652633752*	0,007778*
A-I	0,0000000000000022924995236*	0,008889*
A-J	0,0000000000000188135063098*	0,010000*
B-C	0,0000000000001535711557939*	0,011111*
B-D	0,0000000000014621097665923*	0,012222*
B-E	0,0000000000077998787428868*	0,013333*
B-F	0,0000000000237520274559699*	0,014444*
B-G	0,0000000001043824651159040*	0,015556*
B-H	0,0000000003921834973441560*	0,016667*
B-I	0,0000000018187403394431800*	0,017778*
B-J	0,0000000109608980246589000*	0,018889*
C-D	0,0000000265060956192720000*	0,020000*
C-E	0,0000000930611820035843000*	0,021111*
C-F	0,0000002857406433793840000*	0,022222*
C-G	0,0000009414849929322290000*	0,023333*
C-H	0,0000032865506035822900000*	0,024444*
C-I	0,0000109784351513302000000*	0,025556*
C-J	0,0000303153649629655000000*	0,026667*
D-E	0,0000696833015214867000000*	0,027778*
D-F	0,0001510747976395890000000*	0,028889*
D-G	0,0003479508671506970000000*	0,030000*
D-H	0,0007527678385681180000000*	0,031111*
D-I	0,0014565421847827400000000*	0,032222*
D-J	0,0026145052203776200000000*	0,033333*
E-F	0,0045691768392574000000000*	0,034444*
E-G	0,0077106077204198800000000*	0,035556*
E-H	0,0123206492233679000000000*	0,036667*
E-I	0,0187592571301449000000000*	0,037778*
E-J	0,0289516092430585000000000*	0,038889*
F-G	0,0432595198034065000000000*	0,040000
F-H	0,0628709259068479000000000	0,041111
F-I	0,0908828572543007000000000	0,042222
F-J	0,1278447198176540000000000	0,043333
G-H	0,1771375295143160000000000	0,044444
G-I	0,2525097363417940000000000	0,045556
G-J	0,3507193189482940000000000	0,046667
H-I	0,4724441968684370000000000	0,047778
H-J	0,6333766754534350000000000	0,048889
I-J	0,8143350242746520000000000	0,050000

* Gruplar arasında istatistiksel olarak farklılık olduğunu göstermektedir.

Çizelge 4.9. 10 grup ve n=200 birimlik örneklem için elde edilen p ve FDR değerleri

Karşılaştırılan Gruplar	**t testi p_i değeri**	**FDR testi p_i değeri**
A-B	0,0000000000000000000000000*	0,001111*
A-C	0,0000000000000000000000000*	0,002222*
A-D	0,0000000000000000000000000*	0,003333*
A-E	0,0000000000000000000000000*	0,004444*
A-F	0,0000000000000000000000000*	0,005556*
A-G	0,0000000000000000000000000*	0,006667*
A-H	0,0000000000000000000000000*	0,007778*
A-I	0,0000000000000000000000000*	0,008889*
A-J	0,0000000000000000000000000*	0,010000*
B-C	0,0000000000000000000000000*	0,011111*
B-D	0,0000000000000000000000000*	0,012222*
B-E	0,0000000000000000000000000*	0,013333*
B-F	0,0000000000000000000000000*	0,014444*
B-G	0,0000000000000000000000000*	0,015556*
B-H	0,0000000000000000002220446*	0,016667*
B-I	0,0000000000000000124344979*	0,017778*
B-J	0,0000000000000001794120408*	0,018889*
C-D	0,0000000000000500761654365*	0,020000*
C-E	0,0000000000001165127994085*	0,021111*
C-F	0,0000000000039925138750618*	0,022222*
C-G	0,0000000000280903342808614*	0,023333*
C-H	0,0000000002352973988983820*	0,024444*
C-I	0,0000000019056126101535400*	0,025556*
C-J	0,0000000099129453947632800*	0,026667*
D-E	0,0000000619835252937406000*	0,027778*
D-F	0,0000002226417140480660000*	0,028889*
D-G	0,0000007486843286350630000*	0,030000*
D-H	0,0000027270789359268300000*	0,031111*
D-I	0,0000097220594839388900000*	0,032222*
D-J	0,0000337464786794413000000*	0,033333*
E-F	0,0000877645436575134000000*	0,034444*
E-G	0,0002149456803242940000000*	0,035556*
E-H	0,0004796718525772590000000*	0,036667*
E-I	0,0010091998396619000000000*	0,037778*
E-J	0,0021863971447260500000000*	0,038889*
F-G	0,0047072595860724700000000*	0,040000*
F-H	0,0084841536240105000000000*	0,041111*
F-I	0,0163538642442905000000000*	0,042222*
F-J	0,0289740982818153000000000*	0,043333*
G-H	0,0490194299411033000000000*	0,044444
G-I	0,0828179428948482000000000	0,045556
G-J	0,1445185201491050000000000	0,046667
H-I	0,2411097454516110000000000	0,047778
H-J	0,4199678196323640000000000	0,048889
I-J	0,6828134153085580000000000	0,050000

* Gruplar arasında istatistiksel olarak farklılık olduğunu göstermektedir.

Örneklem hacminin 50 birim olduğu durumda 10 grup için yapılan 45 ikili karşılaştırma sonuçlarının yer aldığı Çizelge 4.7 incelendiğinde, yapılan ikili karşılaştırmalardan t testine göre on dördünün kabul edildiği, diğerlerinin red edildiği görülmektedir. Aynı şekilde FDR değerlerine bakıldığında karşılaştırmalardan on beş tanesinin kabul, diğerlerinin red edildiği görülmektedir.

Örneklem hacminin 100 birim olduğu durumda 10 grup için yapılan 45 karşılaştırma sonuçlarının yer aldığı Çizelge 4.8 incelendiğinde, yapılan ikili karşılaştırmalardan t testine göre dokuzunun kabul edildiği, FDR'ye göre on tanesinin kabul edildiği görülmektedir. Aynı şekilde 3 ve 5 grup karşılaştırmalarında olduğu gibi burada da örneklem hacmi artarken ikili karşılaştırma p değerleri daha da küçülmekte fakat FDR değerleri sabit kalmaktadır.

Örneklem hacminin 200 birim olduğu durumda 10 grup için yapılan 45 ikili karşılaştırma sonuçlarının yer aldığı Çizelge 4.9 incelendiğinde, yapılan ikili karşılaştırmalardan beşinin kabul edildiği, diğerlerinin red edildiği, FDR değerlerine bakıldığında ise beş tanesinin kabul edildiği görülmektedir. Aynı zamanda örneklem sayısı arttıkça kabul edilen hipotez sayılarında bir azalma olmaktadır.

5. TARTIŞMA VE SONUÇ

İkiden fazla grup ortalaması arasındaki farkın anlamlılığının test edilmesinde Varyans Analizinin kullanılabilmesi için bilindiği üzere bazı varsayımların sağlanması gerekmektedir. Bu varsayımlar sağlandığında ve Varyans Analizi uygulandığında sıfır hipotezi red ediliyor ise, ele alınan gruplardan hangileri arasındaki farkın istatistiksel açıdan önemli olduğunun sınanması için çoklu karşılaştırma testlerinden yararlanılmaktadır. Farklı durumlar için kullanılabilecek birçok çoklu karşılaştırma ve aralık testi olmasına rağmen, uygulamada kullanılan bütünleşik F testinde her bir grup ortalaması arasındaki farkın anlamlılığı sınanırken sabit bir α anlamlılık seviyesinin kullanılmasında oluşacak sıkıntı üzerine birçok çalışma mevcuttur. Deneysel ortak hata oranını doğru belirlemek daha sağlıklı sonuçlar elde etme açısından oldukça önemlidir. Mevcut klasik yaklaşımların hiçbiri yanlış bulgu oranı ile ilgilenmemektedir. Bu çalışmanın ana konusu olan yanlış bulgu oranı (FDR) testi bu bağlamda büyük bir önem teşkil etmektedir.

Çalışmadan elde edilen bulgular incelendiğinde, karşılaştırma sayısı ve örneklem hacmi arttıkça kabul edilen sıfır hipotezi sayısındaki azalma dikkat çekmektedir. Çalışmada üç beş ve on grup olması durumunda 50, 100 ve 200 birimlik örneklemler için yapılan ikili karşılaştırmalar göz önüne alınacak olursa 50 birimlik bir örneklem hacmi için 10 grup söz konusu olduğunda yapılan 45 ikili karşılaştırmadan 14 tanesi için sıfır hipotezi kabul edilirken, örneklem hacmi 100 birim olduğunda 45 karşılaştırmadan 9'u, örneklem hacmi 200 olduğunda ise yine 45 ikili karşılaştırmadan 5'i için sıfır hipotezi kabul edilmiştir. Oransal olarak bu değerler 50, 100 ve 200 birimlik örneklemler için sırasıyla %31, %20 ve %11'dir.

Beş grup söz konusu olduğunda yapılan ikili karşılaştırmalar için elde edilen sonuçlar değerlendirilecek olursa, örneklem hacminin 50 olması durumunda yapılan 10 karşılaştırma için kabul edilen sıfır hipotezi sayısı 5 iken 100 ve 200 birimlik örneklemler için bu sayı sırasıyla 3 ve 2'dir. Oransal olarak bu değerler incelendiğinde 50, 100 ve 200 birimlik örneklemler için sırasıyla %50, %30 ve %20'dir.

Benzer şekilde üç grup söz konusu iken yapılabilecek üç ikili karşılaştırma için örneklem hacmi 50 birim olarak ele alındığında, kabul edilen sıfır hipotezi sayısı 2 iken örneklem hacminin 100 ve 200 birime çıkarılmasıyla bu sayı 1'e inmekte, ancak iki yüz birimlik örneklem hacmi için ilgili p değerinin daha küçük olduğu görülmektedir.

t testi ikili karşılaştırma ve FDR testi sonuçları incelendiğinde dikkat çeken bir diğer sonuç ise karşılaştırılacak grup sayısı aynı olmasına rağmen örneklem hacminin artmasıyla FDR değerleri aynı kalmakta ancak ikili karşılaştırma t testi sonucu elde edilen p değeri gittikçe azalmaktadır.

Karşılaştırılacak grup sayısının artmasıyla birlikte aynı örneklem hacimleri için t testi ikili karşılaştırma sonucunda elde edilen p değerinin gittikçe küçük değerler alması, dolayısı ile gruplar arasında istatistiksel açıdan anlamlı bir fark olmadığını ifade eden sıfır ön savının reddine ilişkin olasılığın artması da elde edilen önemli bir diğer sonuç olarak söylenebilir.

Yapılan simülasyon çalışması sonuçlarından da görüldüğü gibi t testine ait ikili karşılaştırma p değerleri örneklem büyüklüğü ve karşılaştırılacak grup sayısı artıkça daha küçük değerler almasına rağmen bir yerden sonra büyüyerek anlamlılık seviyesini koruyamamaktadır. Fakat FDR değerlerine bakıldığında aynı grup sayısında örneklem büyüklüğü artmasına rağmen aynı kaldığı ve anlamlılık seviyesini aşmadığı görülmektedir. Aynı şekilde FDR testinin grup sayısı arttıkça birinci tip hata yapma olasılığının azaldığı, diğer testlere göre daha hassas olduğu sonucuna varılmaktadır.

Literatürde farklı durumlar için kullanılan birçok karşılaştırma testi arasından hangisinin kullanılacağına karar verilmesi oldukça güç bir seçimdir. Birçok varsayımın yanı sıra örneklem büyüklüğü ve grup sayısı kriterleri göz önüne alındığında FDR testinin fazla grup sayısı söz konusu olduğunda deneysel ortak hata oranını kullanan ideal bir yöntem olduğu ifade edilmektedir. Bundan sonra yapılacak çalışmalarda sabit bir α anlamlılık seviyesinin kullanılması yerine deneysel ortak hata oranının doğru belirlenip bahsedilen yöntemler içinden ideal koşullarda hangi yöntemin kullanılmasının uygun olacağı tespit edilmelidir.

KAYNAKLAR

Atil, H. ve Ünver, Y. (2001). Multiple comparisons. *Journal of Biological Sciences,* **1:** 723-727.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society,* **57:** 289-300.

Benjamini, Y. and Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics,* **25:** 60–83.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics,* **29:** 1165–1188.

Carmer, S.G. and Walker, W.M. (1985). Pairwise multiple comparisons of treatment means in argonomic research. *Journal of Argonomic Education,* **14:** 19-26.

Cribbie, R.A. and Keselman, H.J. (2003). Pairwise multiple comparisons: A model comparison approach versus stepwise procedures. *British Journal of Mathematical and Statistical Psychology,* **56:** 167-182.

Cyril, D., Philippe, B. and Thierry, M. (2005). A simple procedure for estimating the false discovery rate. *Bioinformatics,* **21:** 660–668. doi:10.1093 /bioinformatics/bti063

Çömlekçi, N. (2003). Deney Tasarımı İlke ve Teknikleri. Alfa Basım Yayın Dağıtım, İstanbul.

Demirhan, H., Dolgun, N.A., Demirhan, Y.P. ve Dolgun, M.Ö. (2011). Performans of some multiple comparison tests under heteroscedasticity and dependency. *Journal of Statistical Computation and Simulation,* **80:** 1083-1100.

Doğan, İ. ve Doğan, N. (2013). Birinci tür hata'nın kontrolünde yanlış bulgu oranı (False Discovery Rate) yaklaşımı. *Düzce Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi,* (Basım Aşamasında).

Doğan, N. Ve Doğan İ. (2013). Birinci tür hata'nın kontrolü ve adımsal (stepwise) çoklu karşılaştırma testleri. *Düzce Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi*, (Basım Aşamasında).

Dunnett, C.W. and Tamhane, A.C. (1992). A step-up multiple test procedure. *Journal of the American Statistical Association*, **87:** 162-170.

Duncan, D.B. (1955). Multiple range and multiple F tests. *Biometrics*, **11:** 1-42.

Gurarie E. (2008). Introduction to Analysis of Variance. University of Washington, Seattle, QERM 598- Lecture 3.

Hochberg, Y. and Tamhane, A. (1987). Multiple Comparison Procedures: Wiley.

Hochberg, Y. (1988). A sharper bonferroni procedure for multiple tests of significance. *Biometrika*, **75:** 800-803.

Holland, B.S. and Copenhaver, M.D. (1987). An improved sequentially rejective Bonferroni test procedure. *Biometrics*, **43:** 417–423.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**: 65-70.

Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, **75**: 383-386.

İngersoll G.M. (2010). Multiple comparisons of means- a practical guide. *İnernational Journal for Research in Education*, No: 28.

İrgüren, M. (2001). Sağlık Bilimlerinde Varyans Analizi ile İlgili bir Deneme. Yüksek Lisans Tezi, Fırat Üniversitesi, Sağlık Bilimleri Enstitüsü, Elazığ.

Kayri, M. (2009). Araştırmalarda gruplar arası farkın belirlenmesine yönelik çoklu karşılaştırma (Post-Hoc) teknikleri. *Fırat Üniversitesi Sosyal Bilimler Dergisi,* cilt **19:** 51-64.

Keselman, H.J. and Rogan, J.C. (1978). A comparison of the modified- Tukey and Scheffé methods of multiple comparisons for pairwise contrasts. *Journal of the American Statistical Association,* 73: 47-52.

Mendeş, M. (2002). Varyansaların heterojen olması durumunda K istatistiği (KANOVA) ile ANOVA F testinin gerçekleşen 1. tip hata olasılıkları bakımından karşılaştırılması. *Tarım Bilimleri Dergisi,* **8:** 238-241.

Nichols, T. and Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Metods in Medical Research,* **12**: 419-446.

Polat, G. (2006). Hazır Giyim İşletmesinde Yüksek Verimlilik için Varyans analizi ve Barkod Uygulaması. Yüksek Lisans Tezi, Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Kocaeli.

Rao, C.V. and Swarupchand, U. (2009). Multiple comparison prosedures- a note and a bibliography. *Journal of Statistics,* **16:** 66-109.

Scheid, S. and Spang, R. (2003). A False Discovery Rate Approach to Separate the Score Distributions of Induced and Non-induced Genes Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003) March 20–22, Vienna, Austria ISSN 1609-395X.

Shaffer, J.P. (1986). Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association,* **81**: 826-831.

Simes, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance, *Biometrika*, **73:** 751-4.

Soriç, B. (1989). Statistical "Discoveries" and effect-size estimation. *Journal of The American Statistical Association,* **84:** 608-610.

SPSS (2011). IBM SPSS *Statistics* - Base 20.

Storey, J.D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society*, **64**: 479–498.

Storey, J.D., Taylor, J.E. and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society,* **66**: 187–205.

Şenoğlu B. ve Acıtaş Ş. (2010). İstatistiksel Deney Tasarımı: Sabit Etkiler Modeli. Nobel Yayın Dağıtım, 1. Basım, Ankara.

Tukey, J.W. (1949). Comparıng Indıvıdual Means In The Analysıs Of Varıance, *Biometrics*, **5:** 99-114.

Ural A. ve Kılıç İ. (2011). Bilimsel Araştırma Süreci ve SPSS ile Veri Analizi. Detay Yayıncılık, 3. Baskı, Ankara.

Üçkardeş F. (2006). İstatistik Testler Üzerine Bir Çalışma. Yüksek Lisans Tezi, Kahramanmaraş Sütçü İmam Üniversitesi, Fen Bilimleri Enstitüsü, Kahramanmaraş.

Wright, S.P. (1992). Adjusted P- values for simultaneous ınference. *Biometrics,* **48**: 1005-1013.

Yudi, P., Stefan M., Serge K., Arief G. and Alexander P. (2005). False discovery rate, sensitivity and sample size for microarray studies, Bioinformatics, **21:** 3017–3024.

İnternet Kaynakları

1. http://www.stat.cmu.edu/~genovese/ (Christopher R. Genovese, A Tutorial on False Discovery Control)

ÖZGEÇMİŞ

Adı Soyadı	: Aybüke KOCA
Doğum Yeri ve Tarihi	: Zonguldak, 14/09/1987
Yabancı Dili	: İngilizce
İletişim (Telefon/e-posta)	: aybukekoca@hotmail.com

Eğitim Durumu (Kurum ve Yıl)

Lise	: Kozlu Lisesi
Lisans	: Afyon Kocatepe Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü
Yüksek Lisans	: Afyon Kocatepe Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü

EKLER

Ek 1. Veri Türetimine ait kodlar

```
clear all

N=10000; % anakutle hacmi


muA=25; ssA=5.5;
muB=26; ssB=6;
muC=30; ssC=10;

X=muA+ssA.*randn(1,N)';
Y=muB+ssB.*randn(1,N)';
Z=muC+ssC.*randn(1,N)';

veri=[X Y Z];
```

Ek 2. Tekniklere ait kodlar

```
function [ist1 ist2]=teknikler(veri,n)

ctr=0;
dongu=1;
while dongu==1
    rastindis=round((size(veri,1)-1)*rand(n,1))+1;
    secilen=veri(rastindis,:);
    kactane=0;
    for i=1:n
        secx=secilen(i,2);
        sec2=secilen(:,2);
        [z1,z2]=find(secx==sec2);
        if size(z1,1)~=1
            kactane=kactane+1;
        end
    end

    if kactane==0
        secilen=veri(rastindis,:);
        dongu=0;
    else
        ctr=ctr+1;
    end

end


X=secilen(:,1);
Y=secilen(:,2);
Z=secilen(:,3);

G1=[X,X,Y]';
G2=[Y,Z,Z]';

P = mattest(G1,G2);
Q=0.05;
q=Q;

%rejected hipotez sayısı

[n,m]=size(P);
R=zeros(n,1);
fdr_line=[1:m]*q/m;
for i=1:n
    in_fdr=find((P(i,:)-fdr_line)<=0,1,'last');
    if in_fdr>0
        R(i)=in_fdr;
    elseif isempty(in_fdr)==1
```

```
        R(i)=0;
    else
        R(i)=m;
    end
end
Rj=sum(R);
effective_N=Rj;


% Compute p-values passing the fdr using the Benjamini-Hochberg
procedure.
%
% Inputs:
% P - the vector of p-values
% Q - the FDR cutoff
% effective_N - a possible 'true' number of hypothesis we should
normalize
% by (could be bigger then the number of p-values we input since we
didn't
% compute all the p-vals
%
% The output is:
% num_rejected - the # of rejected hypothesis
% fdr_vec - a vector of the fdr cutoff values used by the procedure
(why is this needed?)
% idx - the indices of the rejected hypothesis


if size(P,1)==1
    P=P';
end
Ng=length(P);
if(nargin == 2)
    effective_N = Ng;
end

[sorted_P, idx] = sort(P);
slope = Q ./ effective_N; %%% Ng; % can be bigger
fdr_num_vec = [1:Ng]' * slope;
fdr_num = find(fdr_num_vec > sorted_P);
num_rejected = max(fdr_num);
idx=idx(1:num_rejected);
fdr_vec = Q.*[1:Ng]/Ng;
if(isempty(num_rejected))
    num_rejected = 0;
end
FDR=fdr_vec';

k=sort(P);
ist1=[k];
ist2=[FDR];

sonuc=[ist1 ist2];
```

Ek 3. Simülasyon kodları

```
clear all
load('veri.mat')

n=200;

dx=100000;
itr_sayi=dx/n;

fprintf('\n\nSinan Saracli ---- Program basladi')
fprintf('\nIterasyon no:.')
kat=1;
tablofull=[];

for itr=1:itr_sayi

    if itr==500*kat
        fprintf('.%d',itr)
        kat=kat+1;
    end

    [ist1 ist2]=teknikler(veri,n);

        tablofull(itr,:)=[ist1' ist2'];
    end


ort=mean(tablofull);

fprintf('\n\nProgram basariyla tamamlandi...\n')
```