

## Student t-Testi ve Varyans Analizi Tekniğinde Testin Gücü ve Uygun Örnek Genişliğinin Saptanması

Ensar BAŞPINAR<sup>1</sup>

Handan ÇAMDEVİREN<sup>1</sup>

Fikret GÜRBÜZ<sup>1</sup>

Geliş Tarihi : 15.11.1999

**Özet:** Bu çalışmada, iki ve üç grup ortalamasının karşılaştırılmasında kullanılan Student t-Testi ve Varyans Analizi Tekniğinde istenen güce sahip sonuçların elde edilebilmesi için, gerekli örnek genişliğini tespit etmek amacıyla bir simülasyon çalışması yapılmıştır. Bu simülasyonda, varyansları birbirine eşit, ortalamaları birbirinden farklı olan 6 adet normal populasyon üretilmiş ve iki farklı yaklaşımla bu populasyonlardan örnekler çekilmiştir. Bu iki farklı yaklaşım; a) üretilen 6 populasyonun mümkün olan ikili kombinasyonlarından yararlanılarak, her defasında bu ikili kombinasyonların birinden birer örnek alınarak t-Testi ile karşılaştırılmış ve bu işlem 10 000 defa tekrarlanarak gerçekleşen güç değeri ampirik olarak, çekilen örneklerde eşit ve farklı sayıda gözlem bulunmasına göre çeşitli örnek genişliklerinde hesaplanmıştır. Bu işlemler sonucunda örneklerde eşit sayıda gözlem bulunmasının t-Testinin istenen güce ulaşmasında daha etkili olduğu görülmüştür. b) yukarıdaki 6 populasyonun her defasında üç tanesini kullanmak üzere, 9 farklı ortalama kombinasyonu oluşturulup, her kombinasyondan eşit ve farklı sayıda gözlem içeren örnekler rastgele seçilip, varyans analizi ile F-Testinin gücü 10 000 deneme sonunda ampirik olarak tespit edilmiştir. Bunun sonucunda ise, 3 ortalamaların sıralanışı bakımından ortada yer alan grupta diğer gruplardan daha az sayıda gözlem bulunması halinde, en az toplam gözlem sayısı ile F-Testinin istenen güce ulaştığı görülmüştür.

**Anahtar Kelimeler ;** I. tip hata, II. tip hata , testin gücü, örnek genişliği, t-testi, varyans analizi

### Power of Student t-Test and ANOVA Technique Determining Suitable Sample Size

**Abstract:** The purpose of this study was to determine the suitable sample size to achieve the necessary power for Student's t-Test and ANOVA, using a simulation model in which six normal populations with equal variances and unequal means were simulated. Samples were taken from these populations using two different approaches namely, a. samples were taken from each possible pair combinations of these six populations and sample means were compared using Student's t-Test. This process was repeated for 10 000 times and the power of the test was calculated empirically on different samples sizes with equal and unequal number of observations. Consequently, it was found that samples having equal number of observations are more effective for reaching the desirable power. b. samples having equal and unequal number of observations were taken randomly from nine different mean combinations using tree of the above mentioned populations each time. These mean combinations were compared 10 000 times using ANOVA. The power of F-Test was calculated empirically then. It was found that the desirable power can be achieved with a few number of total observations, if the middle group mean contains lesser number of observation, than other two groups.

**Key Words ;** Type I error, Type II error, power, sample size, t-test, analysis of variance

#### Giriş

$H_0$  hipotezi doğru iken ret edilirse yanlış bir karar verilmiş, bir hata yapılmış olur. Bu yanlışlık I. Tip hata ( $\alpha$ ) olarak isimlendirilir.  $H_0$  hipotezi yanlışken ( $H_1$  hipotezi doğru), kabul edilmesi de II. Tip hata ( $\beta$ ) olarak adlandırılır. II. Tip hata olasılığı ( $\beta$ )'nin 1'den çıkarılmasıyla bulunan olasılık değeri de testin gücü olarak tanımlanır.  $H_1$  doğru iken, kabul edilme olasılığı testin gücü olduğundan, II. Tip hata yapma ( $H_1$  doğru iken, onu ret etme) olasılığı, aslında bir negatif testin gücüdür (Snedecor ve Cochran 1980, Taylor ve Muller 1995). İstatistik testlerdeki ortak amaç ise, bilinmeyen populasyon parametrelerinin tahmin edilmesidir. Aynı parametrenin tahminlerindeki varyasyon ise uygulamada elde edilen veri tiplerinin ve deneme koşullarının farklılığından ileri gelir. Veri tipine veya kurulan denemeye

uygun bir istatistik metodun kullanılmasında başlangıçta belirlenen I. tip hata olasılığı ve istenen güç her zaman sağlanmayabilir. Bu durumda da seçilen örnek genişliklerinin uygun olarak belirlenmesi gerekmektedir. Eğer, bu koşullar sağlanırsa kullanılan test tutucu olmanın yanında, güvenilir sonuçlar da vermektedir (Wassner 1997, Buning ve Kossler 1997).

Uygulamada en yaygın kullanılan I. tip ( $\alpha$ ) ve II. tip ( $\beta$ ) hata yapma olasılıkları sırasıyla 0.01, 0.05, 0.10 ve 0.20 dir. Testin gücü  $1-\beta$  ile gösterilir ve istenen güç değerleri yukarıdaki I. ve II. Tip hata olasılıkları ile paralel olacak şekilde; 0.99, 0.95, 0.90 ve 0.80 dir. Pratikte  $\alpha=0.05$  ve  $1-\beta=0.90$  değerleri test için en yaygın kullanılan değerlerdir (Andrés ve del Castillo 1990, O'Gorman 1990, Gillett 1996, Banik ve ark. 1986).

<sup>1</sup> Ankara Üniv. Ziraat Fak. Zootekni Bölümü Biyometri ve Genetik A.B.D. - Ankara

Deneme planlamasının istatistik aşaması aşağıdaki adımlardan oluşur.

- $H_0$  hipotezinin tanımlanması ve buna bağlı olarak istatistik testin seçilmesi,
- I. ve II. tip hata olasılıklarının belirlenmesi,
- Buna bağlı olarak önem düzeyi ( $\alpha$ ) ve gücün ortaya konması,
- Yanılma olasılığı ve II. tip hata olasılığı dikkate alınarak gerekli örnek genişliğinin hesaplanması.

Eğer, gerekli olan örnek genişliği pratikte temin edilemiyorsa yukarıdaki adımların yeniden tekrarlanması gerekir (Sokal ve Rohlf 1995, Snedecor ve Cochran 1980).

Son yıllarda, araştırmacılar, uygulanan test sonucunda elde edilen sonuçların daha güvenilir olması için testin gücü hakkında bilgi sahibi olmak istemektedirler. Fakat ne yazık ki Tarımsal ve Biyolojik konularda yapılan analizler sonunda uygulanan analiz tekniğinin gücü hakkında yapılan çalışma sayısı son derece azdır. Başlangıçta belirlenen veya istenen gücü sağlayabilmek için uygun testin seçilmesinin yanı sıra materyalin yani, muamele gruplarına düşecek birey sayısının da uygun belirlenmesi gerekmektedir. Bu nedenle denemenin başında ve sonunda araştırmacı ve/veya istatistikçi denemeyi uygun bir şekilde planlamalı ve istatistik analiz aşamasında yukarıda sayılan adımlara titizlikle uymalıdır (Brownie ve Boos 1994, Mahoney ve Magel 1996, Taneichi ve ark. 1998)

Bu çalışmada eşit varyanslara ( $\sigma_i^2=1, i=1,2,...,6$ ) ve farklı ortalamalara ( $\mu_i = k_i, i=1,2,...,6, k_i=0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$ ) sahip olan populasyonların, çeşitli ikili ve üçlü kombinasyonlarından rastgele seçilen eşit ve farklı sayıda deney ünitesi içeren örneklerin karşılaştırılmasında kullanılan testin gücü araştırılmıştır. Bu güç değerlerinden yararlanarak t-testi ve varyans analizi teknikleri için uygun örnek genişliklerinin tespit edilmesi amaçlanmıştır.

#### Materyal ve Yöntem

Araştırmada, simülasyon ile üretilen populasyonlar parametrelerine göre sırasıyla  $N(0.0,1)$ ,  $N(0.5,1)$ ,  $N(1.0,1)$ ,  $N(1.5,1)$ ,  $N(2.0,1)$ ,  $N(2.5,1)$  ve  $N(3.0,1)$  olan ve teorik olarak normal dağılan 6 farklı populasyondur. Bu populasyon ortalamaları arasındaki farklar ( $\Delta$ ) ise sırasıyla  $\Delta_1=0.5$ ,  $\Delta_2=1.0$ ,  $\Delta_3=1.5$ ,  $\Delta_4=2.0$ ,  $\Delta_5=2.5$  ve  $\Delta_6=3.0$  standart sapma olacak şekilde, bütün populasyonların ikili kombinasyonlarından oluşmaktadır. Bu farkların elde edildiği populasyonların her birinden 1'er örnek rastgele çekilmiştir. Çekilen bu iki örneğin genişlikleri en az 2 en çok 100 gözlemden oluşturulmuştur. Bu iki örneğin örnek genişlikleri de eşit ve farklı olmak üzere iki farklı kombinasyonda ele alınıp, ortalamalar arasındaki farklar ( $d_i$ ), Student t-testi ile karşılaştırılıp, gerçekleşen güç değerleri ampirik olarak elde edilmiştir.

Ayrıca, yukarıdaki 6 populasyondan her defasında 3 tanesini kullanarak, eşit ve farklı gözlem içeren örnek kombinasyonları oluşturulmuştur. Bu üç örnek ortalamaları

arasındaki farklar Varyans Analizi tekniği ile karşılaştırılıp, yapılan testin güç değerleri ampirik olarak elde edilmiştir. Üç grup için, belirlenen populasyon ortalamaları kombinasyonları;  $\mu_1; \mu_2; \mu_3$  sırasıyla a) 0.0;0.5;1.0, b) 0.0;0.5;1.5, c) 0.0;1.0;1.5, d) 0.0;0.5;2.0, e) 0.0;1.0;2.0, f) 0.0;1.0;3.0, g) 0.0;1.5;2.0, h) 0.0;1.5;3.0 ve i) 0.0;2.0;3.0 şeklinde olup, bu populasyonlardan çekilen her bir örnekteki gözlem sayısı da, populasyonların birbirine olan uzaklıklarına göre en az 2, en çok 30 olacak şekilde değişik kombinasyonlarda belirlenmiştir.

Belirlenen her bir kombinasyon için 10 000 simülasyon denemesi yapılmış ve her bir denemede üretilen tesadüf sayıları gözlem değeri olarak kullanılmıştır. Tesadüf sayılarının üretilmesinde, Microsoft FORTRAN Power Station'da Fortran 90 Programlama dilinden yararlanılmıştır.

Varyansları eşit, ortalamaları arasında en az 0.5, en fazla 3 standart sapma fark gösteren sırasıyla iki ve üç farklı populasyondan tesadüfen çekilen örnek ortalamaları arasındaki farklar, parametrik testlerden Student t-testi ve F-testi ile kontrol edilmiştir. Bu şekilde hesaplanan t ve F istatistikleri merkezi olmayan dağılımlar göstermektedirler. Bu çalışmada 10 000 simülasyon sonucunda elde edilen t ve F değerlerinin ret edilme olasılıklarından yararlanılarak söz konusu testlerin güç değerleri ampirik olarak belirlenmiştir. Ancak, yukarıda değinilen merkezi olmayan dağılımların eklemeli dağılım fonksiyonları ve yoğunluk fonksiyonlarından yararlanılarak cebirsel olarak da testin gücü hesaplanabilmektedir (Taylor and Muller 1995). Bu çalışmada, ortalamaları bakımından birbirinden farklı olan populasyonların, ortalamaları arasındaki farkların ( $\Delta_i$ ) da birbirinden farklı olacağından hareket edilmiştir. Her bir örnek genişliği ve her bir ortalama kombinasyonunda üretilen tesadüf sayıları için 10 000 deneme yapılmıştır. Denemeler sonucunda ise ret edilen hipotez sayıları ve buna bağlı olarak ret edilme olasılıkları hesaplanmıştır. Bu değerler, yukarıdaki varsayım doğrultusunda testin gücü olarak belirlenmiştir. Elde edilen ampirik sonuçlar, aynı zamanda Minitab (ver. 12.0), Sample Power ve Sigmatat istatistik paket programlarındaki hesaplama sonucunda bulunan güç değerleri ile karşılaştırılarak aynı olup olmadıklarına bakılmıştır.

İstatistik testlerin uygulanabilmesi için birtakım ön şartların yerine getirilmesi gerekir. Bu ön şartların tamamen sağlanamadığı durumlarda uygulanacak test sonucu varılacak kararda, hata yapma olasılığı da artmaktadır. Bilindiği üzere istatistik testlerin hepsinde yanılma payı vardır. Bu pay, test yapmadan önce üzerinde durulan konuya bağlı olarak genellikle %0.1, %1, %5, %10 ve %20 olarak kararlaştırılmaktadır. Biyolojik bilimlerde bu olasılıklar içinde en yaygın kullanılanlar %1 ve %5 tir. Bu yanılma payının büyüklüğü, üzerinde durulan özelliğin ölçü birimine, seçilen testin türüne, uygulanacak teste ait ön şartların yerine gelip gelmemesine, araştırmacının bilgisine vs. göre artar veya azalır. Hata yapma ihtimali olan bu testlerden elde edilen sonuçların güvenilir olması için, testin gücünün yüksek olması gerekir. Testin gücünü artıran şartlar ise şöyle sıralanabilir.

- Uygulanacak teste ilişkin ön şartların sağlanması,
- Test istatistiğinin gerçek dağılımının bilinmesi,
- Verilerin analizinde, kullanılacak alternatif analiz tekniklerinden hangisinin kullanılacağına isabetle belirlenmesi.

Bu şartlar altında istenilen güce sahip ve hatası düşük olan sonuçlar elde edilebilir. Testin gücü en fazla %100 olabilir. Ancak bu güce sahip sonuçlar elde etmek çoğu durumda mümkün değildir. Pratikte, elde edilen sonuçların güvenilir olması için (çift taraflı hipotez kontrolü ve  $\alpha=0.05$ ) uygulanan testin gücünün genel olarak %80'in üzerinde olması istenir (Lui 1994, Taylor ve Muller 1995, Gillett 1996, Welles 1996, Duchateau ve ark. 1998).

Bu çalışmada, t ve F testlerinin kullanımı esnasında yeterince güçlü sonuçların elde edilmesi için uygun örnek genişliğinin ne olması gerektiği üzerinde durulmuştur. Bunun için, farklı popülasyonlardan alınan örneklerde eşit ve farklı sayıda gözlem içeren örnekler çekilmiş ve güç değerleri bulunarak, gücün istenilen değere ulaştığı durumlardaki örnek genişlikleri dikkate alınmıştır. Yapılan simülasyon çalışması, çeşitli kaynaklarda bahsedilen (Zhang 1994, Taylor ve Muller 1995, Mendes 1998), test istatistiğinin aralık tahmini veya dağılımına ait güç eğrileri yardımıyla gücün tespit edilmesine alternatif bir yaklaşım olarak düşünülmüştür. Ayrıca bu yaklaşım, güç eğrilerinden farklı olarak, gruplardaki (örneklerdeki) deney ünitesi sayısının eşit olmadığı durumlarda da testin gücünü hesaplamaya olanak sağlamaktadır.

### Bulgular ve Tartışma

Simülasyon çalışması sonunda elde edilen güç değerleri çekilen örnek sayısı, örnek genişliği ve popülasyon ortalamaları arasındaki fark değerlerine göre Çizelge 1 ve Çizelge 2'de toplu olarak verilmiştir. Elde edilen sonuçlar, örneklerin çekildiği popülasyonların iki (t-Testi) veya üç (F-Testi) tane olmasına göre alt başlıklar halinde şöyledir.

#### 1. Student t-testinde güç sonuçları :

Varyansları 1, ortalamaları arasındaki fark 0.5 standart sapma olan iki popülasyon düşünüldüğünde 10 000 deneme sonucunda gruplardaki örnek genişlikleri eşit olduğu durumlarda, uygulanan testin genel olarak istenilen güce ulaşabilmesi için (>% 80) örnek genişliğinin en az her bir örnekte 70'er gözlem (70;70) olması gerektiği sonucuna varılmaktadır (Çizelge 1.).

Gruplardaki deney ünitesi sayısının birbirinden farklı olduğu durumlarda ise, genel olarak istenilen güce ulaşmada daha fazla gözleme ihtiyaç duyulmaktadır. Mesela gruplardaki gözlem sayısı 50;100 (100,50 ile özdeş) (toplam 150 gözlem) olduğu durumda testin gücü

%82 bulunurken gözlem sayısı 70;70 (toplam 140 gözlem) olması halinde testin gücü %83.7' dir. Bu sonuç, uygulanan testte istenilen güce ulaşmak için, gerekli olan minimum örnek genişliğini belirlemede gruplardaki gözlem sayılarının dengeli olması gerektiğini göstermektedir. Örnek genişlikleri daha da artırıldığında, mesela 70;100 olduğunda, güç değeri %89.1, 80;100 olduğunda %90.7 ve 100;100 olduğunda ise %94.3 olarak bulunmuştur (Çizelge 1).

İki grup ortalaması arasındaki fark 1.0 standart sapma olduğunda, gruplardaki örnek genişlikleri daha küçük sayılardan başlatılmıştır. Gruplardaki gözlem sayıları birbirine eşit ve 2;2 iken, güç değeri %10.0 olup oldukça düşük bulunmuştur. Gruplardaki gözlem sayıları 2;30 iken ise güç %26.0'ya yükselmiştir. Ancak bu artış da istenilen düzeydeki artış değildir. Gözlem sayıları 4;30 olduğunda hesaplanan güç %44.1 olup bir önceki kombinasyonda elde edilen güç değerinin yaklaşık iki katıdır. Buna karşılık toplam gözlem sayısındaki artış 2'dir. Bu durum gruplarda bulunan gözlem sayılarındaki dengenin, gücün artışı üzerine etkisini bir kez daha açıklamaktadır. İstenilen güç değerine ulaşıldığı durumdaki en az örnek genişlikleri ise 18;18 olarak belirlenmiştir. Bu durumda, hesaplanan güç değeri ise %83.3 'dür. Örnek genişlikleri 20;30 iken güç %92.2 ve 30;30 örnek genişliğinde ise %96.5 olup oldukça yüksek değerlerdir. Bu durumda 1 standart sapma farkının önemli olduğunu söylerken yapılan t testinin gücünün istenilen düzeye ulaştığı toplam örnek genişliğinin 36 olması gerektiği söylenebilir. Ancak burada tekrar hatırlatmak gerekirse, bu değer gruplardaki örnek genişliklerinin eşit veya birbirine yakın olduğu zaman geçerlidir. Çünkü toplam örnek genişliği 38 olan 8;30 kombinasyonu için elde edilen güç değeri %69.9 olup düşük bir değerdir. Bu durum gruplardaki gözlem sayılarının dengesiz dağıtılmasının bir sonucudur (Çizelge 1).

Ortalamalar arasındaki farkın 1.5 standart sapma olduğu durumda, düşünülen farklı örnek genişliği kombinasyonlarında ise şu sonuçlar elde edilmiştir. Gruplardaki örnek genişliği 2;2 olması halinde testin gücü %14.9, 7;7 olduğunda % 72, 8;8 iken %79.8 ve 15;15 iken de %97.9 olduğu görülmektedir (Çizelge 1). Gruplardaki örnek genişlikleri birbirinden farklı olan 5;15 kombinasyonunda testin gücü %78.4, 6;15 kombinasyonunda %83.2 ve 10;15 kombinasyonunda %94.1 olarak hesaplanmıştır. Bu sonuçlar doğrultusunda istenilen gücü sağlayan en az örnek genişliğinin 16 (8;8) olması gerektiği söylenebilir (Çizelge 1).

Ortalamalar arasındaki fark 2 standart sapma iken, iki grup ortalaması arasındaki farkın önem testinde istenilen güç (% 87.9) 6;6 kombinasyonunda elde edilmiştir. Örnek genişlikleri birbirinden farklı iken istenilen gücün en düşük değeri 5;7 kombinasyonunda %86.9 olarak bulunmuştur (Çizelge 1).

Çizelge 1. Student t-testi sonunda örnek genişliklerine göre tespit edilen güç değerleri (1-β)<sup>1</sup>

İki populasyon ortalaması arasındaki farkın standart sapma cinsinden değeri											
Δ <sub>1</sub> =0.5σ		Δ <sub>2</sub> =1.0σ		Δ <sub>3</sub> =1.5σ		Δ <sub>4</sub> =2.0σ		Δ <sub>5</sub> =2.5σ		Δ <sub>6</sub> =3.0σ	
n <sub>1</sub> ;n <sub>2</sub>	1-β	n <sub>1</sub> ;n <sub>2</sub>	1-β	n <sub>1</sub> ;n <sub>2</sub>	1-β	n <sub>1</sub> ;n <sub>2</sub>	1-β	n <sub>1</sub> ;n <sub>2</sub>	1-β	n <sub>1</sub> ;n <sub>2</sub>	1-β
5;5	10.6	2;2	10.0	2;2	14.9	2;2	21.8	2;2	29.6	2;2	38.4
10;10	18.7	4;4	22.2	3;3	29.5	3;3	46.0	3;3	63.4	3;3	78.2
15;15	26.3	8;8	46.0	4;4	42.7	4;4	66.4	4;4	83.8	4;4	93.8
20;20	34.4	10;10	55.8	5;5	55.9	5;5	79.2	5;5	93.4	5;5	98.5
30;30	47.4	12;12	64.1	6;6	65.3	6;6	87.9	6;6	97.2		
50;50	69.7	14;14	72.8	7;7	72.0	7;7	92.8				
60;60	77.4	16;16	78.0	8;8	79.8	8;8	96.0				
70;70	83.7	18;18	83.3	9;9	84.6						
80;80	88.2	20;20	87.2	10;10	88.3						
100;100	94.3	30;30	96.5	15;15	97.9						
5;10	13.6	2;30	26.0	2;15	46.1	2;5	48.0	2;5	68.0	2;3	60.2
5;30	17.1	4;30	44.1	3;15	60.4	2;10	64.4	2;10	82.6	2;5	81.1
5;100	19.0	6;30	59.3	4;15	70.8	3;10	78.2	3;7	88.5	3;5	92.5
10;100	31.8	8;30	68.9	5;15	78.4	4;10	87.3	3;10	93.7	4;5	96.6
15;30	32.7	10;30	76.5	6;15	83.2	5;7	86.9	4;6	92.4		
15;100	43.8	12;30	81.7	7;10	81.6	5;10	91.8	4;10	97.2		
20;30	40.0	14;30	85.7	7;15	87.6	6;10	95.2	5;7	96.7		
20;100	52.8	16;30	88.2	8;15	90.7	7;10	96.7	5;10	98.9		
25;30	44.0	18;30	90.6	9;15	92.4						
25;40	49.4	20;30	92.2	10;15	94.1						
25;50	52.5										
30;60	60.0										
30;100	65.9										
50;100	82.0										
60;100	85.7										
70;100	89.1										
80;100	90.7										

<sup>1</sup> Örnek genişliklerinin birbirinin tersi (n<sub>1</sub>;n<sub>2</sub>=25;30 ve n<sub>2</sub>;n<sub>1</sub>=30;25 ... gibi) olma durumları özdeş olup, Çizelge'ye birisi alınmıştır.

Bilindiği üzere ortalamalar arasındaki fark büyüdükçe istenen güce daha az sayıda gözlem ile ulaşılabilir. Mesela fark 2.5 standart sapma olduğu zaman istenilen güce en az 4;4, yani toplam 8 gözlem ile ulaşılabilir (%83.8). Bu durumda, gruptaki gözlem sayıları farklı ise, örnek genişlikleri en az 2;10 (güç %82.6) veya 3;7 (güç %88.5) olmaktadır (Çizelge 1).

Ortalamalar arasındaki farkın standart sapma cinsinden değeri 3 olduğu zaman hemen hemen toplam 6 (3;3) gözlem ile istenilen güce (%78.2) ulaşılabilir. Gözlem sayıları 5;5 olduğu durumda ise hesaplanan güç değeri %98.5 olup maksimum değerine oldukça yakındır (Çizelge 1). Çünkü ortalamaları arasında 3 standart sapma fark olan iki populasyon, gerçek durumlarda birbirlerinden tamamen farklı iki ayrı populasyon olarak düşünülebilir.

## 2. Varyans analizinde testin gücü:

Varyansları birbirine eşit, ortalamaları birbirinden farklı olan 3 populasyondan tesadüfen çekilen, eşit ve farklı sayıda gözlem içeren örnekler kullanılarak elde edilen sonuçlar Çizelge 2' de özetlenmiştir.

Üç populasyondan ardışık olarak ikisinin ortalaması arasındaki fark 0.5 standart sapma (0.0;0.5;1.0) olduğu durumda, her bir grupta 20' şer gözlem yer aldığı zaman

testin gücü %79.1, 30' ar gözlem olduğunda %93.5 olarak bulunmuştur. Gruplarda 5;10;15 ve 10;15;5 şeklinde gözlem yer aldığı durumlarda, testin gücü sırasıyla %38.9 ve %33.8 iken gözlem sayıları gruplara 10;5;15 şeklinde dağıtıldığında, hesaplanan testin gücü %53.9'a yükselmiştir. Bu 3 kombinasyon incelendiğinde toplam gözlem sayıları aynı olmasına rağmen, testin gücünde değişimler olmuştur. Bu sonuca göre, gözlem sayılarının farklı olduğu durum için, en küçük ve en büyük ortalamaya sahip gruplarda, ortalamaların büyüklüğü bakımından sıralamada ortada yer alan gruba nazaran daha fazla sayıda gözlem içermesinin testin gücünde artışa sebep olduğu söylenebilir. Üç grupta da eşit olmayan sayıda gözlem yer aldığı zaman, istenilen güce 25;20;30 kombinasyonunda ulaşılmış ve değeri %90.9 olarak hesaplanmıştır. Ancak 20;30;25 ve 30;25;20 kombinasyonlarında da toplam gözlem sayısı 25;20;30 kombinasyonu ile aynı olmasına karşılık, bu kombinasyonlardaki güç değerleri sırasıyla %84.1 ve %87.5 olarak hesaplanmıştır. Güç değerlerinde gözlenen azalma miktarı, sıralamada ortada yer alan grubun en azından diğer grupların birinden daha fazla gözlem içermesinden kaynaklanmaktadır. Üç grupta, ortalama bakımından birbirinden en uzak olanlarda eşit, ortadaki grupta farklı sayıda gözlemin yer aldığı durumda, testin gücü 25;5;25 örnek genişliği kombinasyonu için %87.5 olarak görülmektedir. Testin gücünün, örnek genişliği

kombinasyonunun 30;5;30 olması halinde ise %93.3'e yükseldiği görülmüştür. Bu sonuçlardan görüleceği üzere, birbirinden en fazla 1 standart sapma uzaklıkta bulunan 3 grup ortalaması arasındaki farkın önem testinde kullanılan F-testinin gücünün istenen düzeye ulaşabilmesi için, en az 55 (25;5;25 kombinasyonu) deney ünitesine ihtiyaç duyulmaktadır. Ayrıca toplam 75 gözlem içeren 25;25;25 ve 25;20;30 kombinasyonlarındaki güç değerleri sırasıyla %88.3 ve %90.9 olarak bulunmuştur. Bu durumda 3 grup için gözlem sayılarının gruplara eşit dağıtılmasından ziyade, ortalamalar bakımından sıralamada tam ortada yer alan gruba, diğer iki gruptan daha az sayıda gözlem ayrılmasının testin gücünü artıracak ve böylece toplam olarak daha az sayıda gözleme ihtiyaç duyulacağı söylenebilir.

Populasyon ortalamalarının standart sapma cinsinden 0.0;0.5;1.5 olduğu kombinasyonda yani, populasyonların ortalamaları arasında en fazla 1.5 standart sapmalı fark bulunduğu, gruplardaki gözlem sayıları eşit ve 10;10;10 yani toplam 30 gözlem var iken, gerçekleşen güç değeri %83' tür. Üç grupta da farklı sayıda gözlem içeren örnek kombinasyonları değerlendirildiğinde, yine toplam 30 gözlem içeren 10;5;15 kombinasyonu için güç değeri %90.1 olarak hesaplanmıştır. Fakat, gözlem sayılarının gruplara dağılışı 5;10;15, 5;15;10 ve 15;10;5 şeklinde olması durumunda gerçekleşen güç değerleri sırasıyla %80.8, %73.8 ve %69.6 olarak hesaplanmıştır. Bu değerlerin daha düşük çıkmasının sebebi ise sıralamada ortada yer alan grubun, diğer iki gruptan en az birinden daha fazla gözlem içermesine bağlıdır. Eğer, 3 gruptaki gözlem sayılarının birbirinden farklı olması gerekiyorsa, ortalaması en büyük olan üçüncü populasyondan çekilen örnekte en çok, en küçük ortalamaya sahip birinci populasyondan çekilen örnekte daha az ve ortada yer alan populasyondan çekilen örnekte ise, en az sayıda gözlemin bulunması halinde yapılacak testte istenen güce daha kolay ulaşabileceği görülmektedir. Ancak böyle bir zorunluluk yok ise, istenen güce ulaşmak için gerekli olan toplam en küçük örnek genişliği, en küçük ve en büyük ortalamaya sahip iki örneğe eşit, ortadaki örneğe ise daha az sayıda gözlem içerecek şekilde dağıtılması önerilebilir. Bu sonuçlar doğrultusunda, populasyonların 0.0;0.5;1.5 ortalama kombinasyonunda, istenen güce ulaşmak için gerekli en az toplam örnek genişliği ve gruplardaki gözlem sayılarının sırasıyla 25 ve 10;5;10 olması gerektiği söylenebilir (Çizelge 2).

Yukarıda bahsedilen açıklamalar genel hatlarıyla 0.0;1.0;1.5, 0.0;0.5;2.0, 0.0;1.0;2.0, 0.0;1.0;3.0, 0.0;1.5;2.0, 0.0;1.5;3.0 ve 0.0;2.0;3.0 ortalama kombinasyonları için de geçerlidir. Yapılan genel açıklamalara ilaveten ortalamalar arasındaki uzaklığın eşit veya farklı olmasının, örnek genişliğini belirlemede önemli bir role sahip olmadığı da söylenebilir. Bu durum 0.0;0.5;2.0 ortalama kombinasyonu ile 0.0;1.0;2.0 ortalama kombinasyonunun birlikte incelenmesiyle de görülmektedir. Her iki durumda da standart sapma cinsinden ortalamalar arasındaki maksimum fark 2 dir. Ancak birinci ortalama kombinasyonunda, birinci ve ikinci

populasyon arasında 0.5 standart sapmalı fark var iken ikinci ile üçüncü populasyon arasında 1.5 standart sapmalı fark vardır. İkinci ortalama kombinasyonunda ise üç populasyon birbirinden eşit uzaklıkta olup, aralarındaki fark 1 standart sapmadır. Her iki durumda da, toplam gözlem sayısı 18 ve bunun gruplara dağılımı da 8;2;8 olduğunda istenen güce ulaşılmaktadır. Sonuçta ortadaki ortalamaların diğer ikisinden eşit uzaklıkta olması veya her hangi birisine daha yakın olmasının, önemli bir etken olmadığı görülmektedir. Yukarıda verilen 7 farklı ortalama kombinasyonu için, istenen güce sahip minimum toplam gözlem sayısı ve örnek gruplarına dağılımları sırasıyla aşağıdaki gibidir (Çizelge 2).

Populasyon ortalamalarının kombinasyonu 0.0;1.0;1.5 olduğunda, gözlem sayılarının gruplara dağılımı 10;5;10 şeklinde ise istenen güce ulaşılmaktadır. Denemeye alınan örnek genişliği kombinasyonlarından 15;15;15 için güç değeri %95.9 olup, oldukça yüksek bir değerdir. Ancak örnek genişliklerinin 15;10;15 olması durumundaki güç değeri (%95.7) hesaplanan en yüksek değerinden önemli derecede farklı olmayıp hemen hemen aynı kabul edilebilir. Fakat son durumda toplam 40 gözleme ihtiyaç duyulurken, diğerinde toplam 45 gözleme ihtiyaç duyulmaktadır. Bu sonuç ise, gruplara gözlem sayısının dağıtılması konusunda yapılan açıklamaları desteklemektedir.

Populasyon ortalamalarının kombinasyonu 0.0;0.5;2.0 olduğu durumda, istenen gücü sağlayan örnek genişliği kombinasyonu 8;2;8 olup, hesaplanan güç değeri %91.5' tir. Yine aynı durumda toplam 18 birey içeren 6;6;6 örnek genişliği kombinasyonu için, bulunan güç değeri %83.8 olup, diğerinden biraz daha düşüktür. Bu durum, gözlem sayılarının gruplara eşit dağıtımındansa, ortadaki gruba daha az sayıda gözlem ayırmanın, hesaplanan güç değerini artıracak olduğunu göstermektedir. Ayrıca bu durumda ele alınan örnek genişliği kombinasyonlarından 10;8;10 ve 10;10;10 için hesaplanan güç değerleri sırasıyla %98.1 ve %98.2 olup en büyük değerine çok yakın çıkmıştır (Çizelge 2).

Populasyon ortalamaları kombinasyonu 0.0;1.0;2.0 iken, elde edilen sonuçlar ortalamaların 0.0;0.5;2.0 olduğu duruma çok yakındır. İstenen gücü sağlayan toplam gözlem sayısı 18 olup, testin gücü, 6;2;10 ve 8;2;8 örnek genişliği kombinasyonlarında sırasıyla %89 ve %90.9 olarak tespit edilmiştir. En yüksek güç değeri ise, 10;10;10 ve 10;8;10 örnek kombinasyonlarında sırasıyla %97.3 ve %97.2 olarak hesaplanmıştır (Çizelge 2).

Populasyon ortalamaları 0.0;1.0;3.0 standart sapma kombinasyonu olarak ele alındığında, en küçük ortalama ile en büyük ortalama arasındaki fark, artık belirgin bir şekilde büyüdüğü için, istenen güce daha az sayıda gözlem ile ulaşılmıştır. Bu nedenle, böyle durumlar için simülasyon denemesinde az sayıda gözlem içeren örnekler kullanılmıştır. Toplam 12 gözlem içeren 4;4;4, 3;4;5, 5;4;3 ve 4;5;3 örnek genişliği kombinasyonlarında gerçekleşen güç değerleri sırasıyla; %90.7, %91.2, %88.2 ve %85.2 olarak tespit edilmiştir (Çizelge 2).



Populasyon ortalamaları 0.0;1.5;2.0 standart sapma kombinasyonu olarak belirlendiği durumda, istenen güce ulaşmak için minimum 17 veya 18 gözleme ihtiyaç durulmaktadır. Bu gözlemlerin gruplara dağılımı ise 6;6;6, 5;6;7, 6;7;5, 7;6;5, 7;5;6, 6;7;6 veya 6;5;6 şeklinde olabilir. Bu kombinasyonlardan 7;5;6 kombinasyonundaki güç değeri %86.8 olup diğerlerinden biraz daha yüksek bulunmuştur. Bu hal için, hesaplanan en yüksek güç değeri %90.1 olup, gözlemlerin gruplara 7;7;7 şeklinde eşit olarak dağıtılması sonucunda elde edilmiştir (Çizelge 2).

Populasyon ortalamaları 0.0;1.5;3.0 ve 0.0;2.0;3.0 standart sapma olduğu hallerde, hesaplanan güç değerleri birbirine yakın çıkmıştır. Her iki durumda da toplam 12 gözlem ile istenen güç değeri elde edilebilmiştir. Ancak bu 12 gözlemin gruplara dağıtımı 5;2;5 şeklinde olduğu zaman, güç değeri daha da yüksek çıkmıştır. Her iki ortalama kombinasyonunda hesaplanan en yüksek güç değeri sırasıyla %96.9 ve %97.2 olup 5;5;5 örnek genişliği kombinasyonundan elde edilmiştir (Çizelge 2).

### Sonuç

Çalışmanın sonunda, Student t-testi ve basit varyans analizi tekniğinin uygulanabilmesi için gerekli ön şartlar sağlandıktan sonra, istenen güce sahip sonuçların elde edilebilmesi için toplam gözlem sayısı ve bunların gruplara dağıtımı ile ilgili olarak, burada yapılan öneriler doğrultusunda, kurulacak bir denemeden elde edilen sonuçlarının güvenilirliğinin yüksek olacağı, gerçekleşen güç değerleri dikkate alınarak ileri sürülebilir. Bu açıklamalar doğrultusunda t-testinin ön şartları gerçekleştiği durumda, yapılan hesaplamalar sonucunda testin gücünün başlangıçta belirlenen seviyeyi koruması için özellikle gruplarda eşit sayıda veya dengeli bir şekilde gözlem bulundurulması gerektiği, aksi taktirde ihtiyaç duyulan toplam gözlem sayısının artacağı söylenebilir. Yani, Student-t testinin uygulanması halinde, gerekli örnek genişlikleri için;

a) çekilen örneklerde eşit sayıda gözlem bulunduğu durumlarda; i)  $\Delta=0.5\sigma$  iken istenilen güç için her bir örnekte en az 70'er gözlem, ii)  $\Delta=1.0\sigma$  iken istenilen güç için her bir örnekte en az 18'er gözlem, iii)  $\Delta=1.5\sigma$  iken istenilen güç için her bir örnekte en az 9'ar gözlem, iv)  $\Delta=2.0\sigma$  iken istenilen güç için her bir örnekte en az 6'şar gözlem, v)  $\Delta=2.5\sigma$  iken istenilen güç için her bir örnekte en az 4'er gözlem ve vi)  $\Delta=3.0\sigma$  iken ise istenilen güç için her bir örnekte en az 4'er gözlem bulunması gerektiği,

b) çekilen örneklerde farklı sayıda gözlem bulunduğu durumlarda; i)  $\Delta=0.5\sigma$  iken istenilen güç için, her bir örnekte en az 60;100 (veya 100;60) gözlem, ii)  $\Delta=1.0\sigma$  iken istenilen güç için, her bir örnekte en az 12;30 gözlem, iii)  $\Delta=1.5\sigma$  iken istenilen güç için, her bir örnekte en az 6;15 gözlem, iv)  $\Delta=2.0\sigma$  iken istenilen güç için, her bir örnekte en az 4;10 gözlem, v)  $\Delta=2.5\sigma$  iken istenilen güç için, her bir örnekte en az 2;10 gözlem ve vi)  $\Delta=3.0\sigma$  iken ise istenilen güç için, her bir örnekte en az 2;5 gözlem bulunması gerektiği söylenebilir.

Üç grup ortalamasının karşılaştırılmasında kullanılan varyans analizi tekniğinde istenen güce ulaşmada, uygun örnek genişlikleri için de, ortalamaların sıralanmasında ortada yer alan grubun, diğer iki gruptan daha az sayıda gözlem içermesinin, testin gücünü artırmada önemli bir etken olduğu sonucuna varılmıştır. Eğer, 3 gruptaki gözlem sayılarının birbirinden farklı olması gerekiyorsa, ortalaması en büyük olan üçüncü populasyondan çekilen örnekte en çok, en küçük ortalamaya sahip birinci populasyondan çekilen örnekte daha az ve ortada yer alan populasyondan çekilen örnekte ise, en az sayıda gözlemin bulunması halinde yapılacak testte istenen güce daha kolay ulaşılabileceğini de söylemek mümkündür. Ancak böyle bir zorunluluk yok ise, istenen güce ulaşmak için gerekli olan toplam en küçük örnek genişliği, en küçük ve en büyük ortalamaya sahip iki örneğe eşit, ortadaki örneğe ise daha az sayıda gözlem içerecek şekilde dağıtılması önerilebilir.

Ayrıca, ortalamalar arasındaki uzaklığın eşit veya farklı olmasının, örnek genişliğini belirlemede önemli bir role sahip olmadığı da söylenebilir. Yani, sıralamada ortadaki ortalamaların diğer ikisinden eşit uzaklıkta olması veya her hangi birisine daha yakın olmasının, örnek genişlikleri üzerinde önemli bir etken olmadığı sonucuna varılmıştır.

### Kaynaklar

- Andrés, A. M. and J. D. L. del Castillo, 1990. Multiple Choice Tests: Power, Length and Optimal Number of Choices per Item. *Brit. Jour. Of Math. And Statist. Psc.* (43), 57-71.
- Banik, N., K. Kohne and P. Bauer, 1986. On the Power of Fisher's Combination Test for 2-Stage Sampling in the Presence of Nuisance Parameters. *Biometrical Jour.* 38(1), 25-37.
- Brownie, C. and D. D. Boos, 1994. Type I Error Robustness of ANOVA and ANOVA on Ranks When the Number of Treatments is Large. *Biometrics.* 50(2), 542-549.
- Buning, H. and W. Kossler, 1997. Power of Some Tests for Umbrella Alternatives in the Multisample Location Problem. Fisher's Combination Test for 2-Stage Sampling in the Presence of Nuisance Parameters. *Biometrical Jour.* 38(1), 25-37.
- Duchateau, L., B. Modermott and G. J. Rowlands. 1998. Power Evaluation of Small Drug and Vaccine Experiments with Binary Outcomes. *Statistics in Medicine*, 17(1), 111-120
- Gillett, R. 1996. Retrospective Power Surveys. *The Statistician.* 45(2), 231-236.
- Lui, K. J., 1994. A Group Sequential Method for One Standard Control and More Than One Experimental Treatment. *Biometrical Journal*, 36(5) 515-529
- Mahoney, M. and R. Magel, 1996. Estimation of the Power of the Kruskal-Wallis Test. *Biometrical Jour.* 38(5) 613-630.
- Mendeş, M., 1998. Parametre Tahmininde ve Ortalamalar Arasındaki Farkın Önem Testlerinde Örnek Genişliğinin Tespiti. *A.Ü. Fen Bilim. Enst. Yük. Lis. Tezi. Basılmamış.* 56 S.
- O' Gorman, T. W., 1995. The effect of Unequal Estimation of the Power of the Kruskal-Wallis Test. *Commun. Statist-Simula.*, 24(4), 853-867.

Snedecor, G. W. and W. G. Cochran. 1980. *Statistical Methods*. Seventh Ed. The Iowa State University Press. Amer, Iowa, U.S.A. 507 S.

Sokal, R. R. and F. J. Rohlf 1995. *Biometry. The Principles and Practice of Statistics in Biological Research*. Third Ed. W.H. Freeman and Company, New York. U.S.A. 887 S.

Taneichi, N. Y. Sekia and H. Imai, 1998. A Power Approximation of the Test of Homogeneity for Multinomial Populations Based on a Normalizing Transformation. *Comm. In Stat.-Theor. and Methods*. 27(7),1677-1691.

Taylor, D. J. and K. E. Muller, 1995. Computing Confidence Bounds for Power and Sample Size of the General Linear Univariate Model. *The American Statistician. Methods*. 49(1), 43- 47.

Wassmer G., 1997. A Technical Note on the Power Determination Fisher's Combination Test. *Bicometrical Jour.* 39(7), 831-838.

Wellek, S., 1996. A New Approach to Equivalence Assessment in Standard Comparative Bioavailability Trials by Means of the Mann-Whitney Statistic. *Biometrical Journal*, 38(6), 695-710.