Uludağ Üniv Tıp Fak Derg 2: 127-130, 1993

İki Ortalama Farkına İlişkin Hipotezlerin z ve t Testi İle Kontrolünde, Verilen Kararların Doğruluk Dereceleri*

Bülent Ediz**, İsmet Kan***

ÖZET. *Biyoistatistikte, kurulan hipotezlerin kontrolü toplum yapısı, örneklerdeki birim sayıları ve toplumlara ilişkin varyansların bilinip, bilinmemelerine göre değişik yollarla yapılmaktadır. Bu araştırmada, yapay olarak oluşturulan ve parametreleri $\mu = 50$, $\sigma = 10$, $N = 2996$ olan normal ve $\mu = 48.734$, $\sigma = 11.36$, çarpıklık katsayısı $Ç_k = 0.1533$, basıklık katsayısı $K_m = -0.136$ ve $N = 3000$ olan anormal, iki toplumdan n_x, n_y (n_x ve n_y 5'ten 50'ye kadar 5 ve 5'in katlarını alacak şekilde) birimli 100'er örnek rastgele seçilerek toplam 10000 örnek incelendi. Her iki toplumda, H_o hipotezinin $\alpha = 0.05$ düzeyinde z ve t'ye göre red oranları, n_x ve n_y'nin değişik kombinasyon gruplarında hesaplanıp birbirleriyle karşılaştırıldılar.*

Anahtar Kelimeler .Ortalama .Fark .Hipotez .z testi .t testi.

The Degrees of Truth of Decisions About Hypotheses of Two Means Differences When Tested by z and t Tests

SUMMARY. *In biostatistics, testing of hypotheses are made by various ways. Choosing the way depends on structure of population, sample sizes are known or not. In this study, two populations were formed artificially. The first population with $\mu = 50$, $\sigma = 10$ $N = 2996$ was normal and the other with $\mu = 48.734$, $\sigma = 11.36$, coefficient of skewness ($Ç_k$) = 0.1533, coefficient of kurtosis (K_m) = − 0.136, N = 3000 was not normal. 100 samples of n_x, n_y (n_x and n_y aries from 5 unit) were drawn randomly from each x and y population. The total samples investigates were 10000. For both population, at the significant level of $\alpha = 0.05$ the rejection ratios of zero hypotheses according to z and t tests were calculated for each combination group of n_x, n_y and comparisons were made among them.*

Key Words .Mean .Difference .Hypothes .z test .t test.

Örneklerin seçildikleri ana kütleler belirli sayıda birimlerden oluşabileceği gibi sonsuz derecede büyük de olabilirler. Biyoistatistiğin temel amacı, parametresi bilinmeyen bir ana kütleden rastgele olarak seçilmiş ve daha az sayıda oluşan örnek birimlerini incelemek suretiyle, ana kütle hakkında genel yargılara varmaktır (ana kütle parametrelerini tahmin etmektir).

Bazı durumlarda, bilinmeyen iki ana kütle ortalamasının karşılaştırılması gereği ortaya çıkmaktadır. Yeni geliştirilen ya da var olan, iki ilaç veya yöntemin iki ayrı gruba uygulanmasından elde edilen sonuçların karşılaştırılması gibi. Bu da tahmin edilen ana kütle parametrelerinin karşılaştırılmasıyla olur. İki ayrı ana kütleden çekilen örnek ortalamaları yardımıyla toplum ortalamalarının aynı olup olmadıkları karşılaştırılır. Örnek ortalamalarının farklılığı iki kaynaktan ileri gelmektedir. Bunlar; deneysel hata ve örneklere uygulanan farklı işlemlerdir. Deney planlaması yolu ile deneysel hata minimuma indirildiğinde, aradaki fark yalnızca işlem farklılığından ortaya çıkmış olacaktır. Ortaya çıkan bu farklılığın,

* U.Ü. Sağlık Bil. Enst. Yüksek Lisans Tezi olarak verilmiştir.
** Araş. Gör.; Uludağ Ü. Tıp Fak. Biyoistatistik BD.
*** Prof. Dr.; Uludağ Ü. Tıp Fak. Biyoistatistik BD.

Geliş Tarihi: 2.11.1992
Kabul Tarihi: 29.3.1993

işlemlerin farklılığından mı yoksa tesadüfi olarak mı oluştuğu hipotez testleri ile ayırt edilir. İki ayrı toplum parametresi farkı için hipotez kontrolü z ve t testleri ile yapılmaktadır. Bu testlerden herhangi birinin seçimi, örneklerdeki birim sayılarına ve toplum variyanslarının bilinip bilinmemesine bağlıdır. Uygulamada, toplum variyansı bilinmediği zaman, örnek birimleri küçük (n , n_y < 30) olduğunda student t testi, örnek birimleri büyük (n , $n_y \geq 30$) olduğunda z testi ve toplum variyansı bilindiğinde de z testi kullanılmaktadır[1-7].

Örneklerin büyüklüğü için seçilen 30 sınır değeri kesin olmayıp yuvarlak olarak alınmaktadır. Merkezi limit teoremine göre bu sayı yukarılara çekildiği sürece örneğe büyüklük vasfı kazandırmaktadır. z testinin uygulanması, t testine göre daha pratik olduğu için onun tercih edilmesi istenir. Ancak bazı durumlarda, koşullar uygun olmadığı için t testinin uygulanması yoluna gidilmektedir. Bu araştırmanın amacı, istatistiğin en önemli kavramları arasında yer alan merkezi limit teoremine göre, normal toplumdan çekilen örneklemlerin dağılımlarının da normal dağılım göstermesi özelliğinin iki ortalama farkına ilişkin dağılımlar için de örneklerin birim sayıları değişirken olup olmadığını kontrol etmektir.

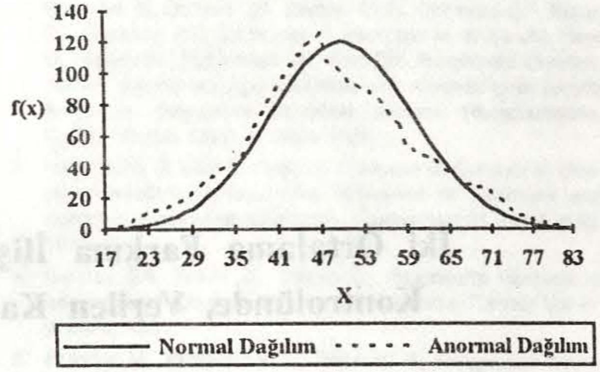
Gereç ve Yöntem

Verileri elde etmek için normal ve normal olmayan iki ayrı toplumdan yararlanıldı. Normal dağılım değerlerini oluşturmak için normal dağılım fonksiyonundan yararlanıldı[1-3,8-15].

Çalışmada normal dağılım olarak $\mu = 50$ ve $\sigma = 10$ parametreli dağılım seçildi ve eldeki imkanlardan (bilgisayar belleği) maximum düzeyde faydalanabilmek için N = 2996 olarak saptandı. Normal dağılım fonksiyonunda N = 2996 alındığında değişken değerlerinin (x'lerin) $17 \leq x \leq 83$ aralığında değiştiği, bu aralığın dışında fonksiyon değerlerinin sıfır (0) olduğu görüldü. Normal olmayan dağılım için normal dağılımdaki frekans değerleri (f(x)'ler) değiştirilerek eğri çarpıtıldı ve N = 3000 veri için $\mu = 48.734$, $\sigma = 11.36$, Ç = 0.1533 ve $K_m = -0.136$ olarak bulundu. Bu dağılımda da değişken değerleri (x'ler) $17 \leq x \leq 83$ aralığında değişmektedir.

Bu iki dağılıma ait frekans poligonu grafiği Şekil:1 de görülmektedir.

Normal ve anormal dağılımlarda aynı değerler, iki ortalama farkına $(\bar{x}-\bar{y})$ ilişkin hipotez testi yapılacağından dolayı, hem x hem de y değerleri olarak alınmıştır. Her iki toplumdan da n ve n_y hacimlerinde 100'er örnek seçilerek bunların ortalamaları arasındaki farka ait sıfır hipotezi (H_0: $\mu = \mu_y$) ve alternatif hipotez (H_1: $\mu \neq \mu_y$) $\alpha = 0.05$ anlamlılık



Şekil: 1
Normal ve anormal dağılımların iç içe gösterimi

düzeyinde hem z ve hem de student t testi ile kontrol edildiler. Başlangıçta z testi varsayımları yerine getirildiği için bu test doğru kabul edildi. Anlamlılık düzeyi $\alpha = 0.05$ kabul edildiği için bu testten beklenen hipotez reddi sayısı 100 testten 5 olacaktır.

Her iki dağılımda da önemli olduğunu tesbit ettiğimiz A (n = 5, n_y = 50'de 100 örnek), B(n = 50, n_y = 50'de 100 örnek), C(n = 5, n_y = 50 veya n = 50, n_y = 5'te 200 örnek), D(n, $n_y \geq 30$'da 2500 örnek), E(n, n_y<30'da 2500 örnek), F(n $\geq$30, n_y< 30 veya n <30 $n_y \geq$ 30'da 5000 örnek) ve G(n_y = n_y'de 1000 örnek) örnek büyüklüklerinde elde etmiş olduğumuz hipotez kabul ve reddi oranlarını kendi içerisinde bağımlı örneklerde z testi ile, değişik örnek büyüklüklerinden elde edilen dağılımları ise Kolmogorov-Smirnov testi ile karşılaştırıldı.

Bulgular ve Sonuçlar

Normal ve anormal (normal olmayan) dağılımlara ait frekans dağılımları arasında fark olup olmadığını anlamak için Kolmogorov-Smirnov testi ile kontrol edildi ve dağılımlarının birbirlerinden farklı olduğu belirlendi (P<0.001).

Her iki testte de iki ortalama farkına ilişkin 4 durum ortaya çıkmıştır. Bunlar;

1. z ve t (veya Z_s) testinin reddi
2. z ve t (veya Z_s) testinin kabulü
3. z kabul t (veya Z_s) red
4. z red t (veya Z_s) kabul'dür.

Değişik örnek büyüklüklerinde (A–G) örneklerin test sonuçlarının oransal dağılımları ve 3.-4. durum oran farklarına ait test sonuçları I.-VI. tablolarda verilmiştir. Burada Z_s örnek variyanslarına göre hesaplanan test istatistiği değeridir.

Tablo: I- Normal dağılımda örnek variyanslarına göre elde edilen hipotez kabulü ve reddi oranları (%)

	1. Durum	2. Durum	3. Durum	4. Durum	3-4'ün Test Sonuçları
A	4.00	93.00	0.00	3.00	$P>0.05$
B	5.00	95.00	0.00	0.00	$P>0.05$
C	3.00	90.00	1.00	6.00	$P<0.01$
D	5.72	93.68	0.04	0.56	$P<0.001$
E	4.48	92.88	0.32	2.32	$P<0.001$
F	4.72	92.90	0.46	1.92	$P<0.001$
G	6.30	92.80	0.00	0.90	$P<0.01$

Tablo: II- Normal dağılımda anakütle variyansına göre elde edilen hipotez kabulü ve reddi oranları (%)

	1. Durum	2. Durum	3. Durum	4. Durum	3-4'ün Test Sonuçları
A	3.00	93.00	4.00	0.00	$P<0.05$
B	5.00	95.00	0.00	0.00	$P>0.05$
C	3.50	95.50	0.50	0.50	$P>0.05$
D	5.28	92.92	1.28	0.52	$P<0.01$
E	4.04	93.32	2.00	0.64	$P<0.001$
F	4.24	93.92	1.24	0.60	$P<0.001$
G	4.90	92.60	2.30	0.20	$P<0.01$

Dağılımları incelerken t testi yerine, örneklerden hesaplanmış variyanslarla Z_s testi kullanılarak, anakütle variyansıyla hesaplanan Z_σ testi ile karşılaştırıldı.

Tablo: III- Normal dağılımda Z_σ ve Z_s'ye göre hipotezlerin kabulü ve reddi oranları (%)

	1. Durum	2. Durum	3. Durum	4. Durum	3-4'ün Test Sonuçları
A	3.00	93.00	0.00	4.00	$P<0.05$
B	5.00	95.00	0.00	0.00	$P>0.05$
C	3.50	90.50	0.50	5.50	$P<0.01$
D	4.88	93.20	0.52	1.40	$P<0.01$
E	3.64	92.16	1.04	3.16	$P<0.001$
F	4.34	92.68	0.70	2.28	$P<0.001$
G	4.90	92.60	0.20	2.30	$P<0.001$

Her iki dağılımda da bütün tablolarda A-G parametrelerinin dağılımları kendi içlerinde Kolmogorov-Smirnov testi ile karşılaştırıldıklarında dağılımlarının biribirlerinden farklı olmadığı görülmüştür ($P>0.05$).

Tablo: IV- Anormal dağılımda örnek variyanslarına göre elde edilen hipotez kabulü ve reddi oranları (%)

	1. Durum	2. Durum	3. Durum	4. Durum	3-4'ün Test Sonuçları
A	6.00	91.00	0.00	3.00	$P>0.05$
B	6.00	94.00	0.00	0.00	$P>0.05$
C	5.00	87.00	1.50	6.50	$P<0.05$
D	5.36	94.12	0.12	0.40	$P>0.05$
E	5.44	91.84	0.32	2.40	$P<0.05$
F	4.82	92.30	0.60	2.28	$P<0.001$
G	5.80	93.50	0.00	0.70	$P<0.01$

Tablo: V- Anormal dağılımda anakütle variyansına göre elde edilen hipotez kabul ve reddi oranları (%)

	1. Durum	2. Durum	3. Durum	4. Durum	3-4'ün Test Sonuçları
A	5.00	91.00	4.00	0.00	$P<0.05$
B	6.00	93.00	0.00	1.00	$P>0.05$
C	7.00	91.50	0.00	1.50	$P>0.05$
D	4.52	93.48	1.32	0.68	$P<0.05$
E	4.96	92.28	2.20	0.56	$P<0.001$
F	4.48	93.42	1.42	0.68	$P<0.001$
G	5.00	92.70	1.50	0.80	$P>0.05$

Tablo: VI- Anormal dağılımda Z_σ ve Z_s'ye göre elde edilen hipotez kabulü ve reddi oranları (%)

	1. Durum	2. Durum	3. Durum	4. Durum	3-4'ün Test Sonuçları
A	5.00	91.00	0.00	4.00	$P<0.05$
B	6.00	93.00	1.00	0.00	$P>0.05$
C	6.50	86.50	2.00	5.00	$P>0.05$
D	4.52	93.48	0.68	1.32	$P<0.05$
E	4.68	91.32	0.84	3.16	$P<0.001$
F	4.24	92.12	0.82	2.82	$P<0.001$
G	5.00	92.70	0.80	1.50	$P>0.05$

Tartışma

W. Gosset, z testi yapılırken $n<30$ olduğu zaman, ana kütle variyansı (σ^2) yerine örneklerden hesaplanan S^2 konulduğunda verilen kararların tam doğru olmadıklarını ve bir hatanın varlığını ortaya koymuştur[1-4,9,16]. Aynı durum bu çalışmada da görülmüştür. n , $n_y<50$ olması halinde z testi yapılırken ana kütle variyansı yerine örneklerden hesaplanan S^2 ve S_y^2 nin kullanılmasıyla bir hata ortaya çıkmaktadır. n , $n_y \geq 50$ olması halinde merkezi limit teoremine göre büyüklük vasfı kazanıldığından dolayı ana kütle variyansı yerine örneklerden hesaplanan variyanslar kullanıldığında bu hata ortadan kalkmaktadır.

n_x, n_y < 50 olması halinde, z testi yapılırken ana kütle variyansı yerine örneklerden hesaplanan variyanslar kullanıldığında, yapılan hata % 7.16, ana kütle variyansı kullanıldığında ise % 4.37 bulunmuştur. Bu iki hata oranı arasındaki fark anlamlıdır ($P<0.001$). Bu sonuçta göstermektedir ki, 50'den küçük örnekler için σ^2 bilinmediğinde $S_x{}^2$ ve $S_y{}^2$ yi ayrı ayrı kullanmak yerine bunların ortak variyansı olan S^2 yi kullanmak daha uygun olacaktır. Çünkü, ortak variyans kullanıldığında hata oranı % 7.16'dan % 6.04'e düşmektedir. Bu düşüş istatistiksel olarak anlamlıdır ($P<0.001$).

n_x, $n_y \geq 50$ olması halinde ise yapılan hata oranları beklenilen düzeyde olup birbirlerinden farklı değildir ($P > 0.05$). n_x, $n_y \geq 30$ durumunda ana kütle variyansı yerine, örneklerden hesaplanan variyanslar konulduğunda z testine göre yapılan hata % 6.28 ve n_x, $n_y < 30$ olması durumunda ise % 6.80'e çıkmaktadır. Örnekler küçüldüğünde hata artmasına karşın anlamlı bir artış değildir ($P>0.05$). Bu da gösteriyor ki, örnek büyüklükleri ne olursa olsun, $\sigma_x{}^2$ ve $\sigma_y{}^2$ nin kullanılmaları en uygun olanıdır. Bunların bilinmemesi halinde ise ortak variyansın kullanılması, ayrı ayrı örnek variyanslarının kullanılmasından daha iyi sonuç vermektedir.

Kan, yaptığı çalışmada tek ortalama farkına ililşkin t testi kullanıldığında, $n>30$ olması halinde yapılan hatayı % 5.20, $n \leq 30$ olması halinde de yapılan hatanın % 6.70 olduğunu ve aralarındaki farkın anlamlı olmadığını bulmuştur[8]. Bu çalışmada, iki ortalama farkı için t testinde yapılan I. tip hata oranları beklenen % 5.00 değerine karşın n_x, $n_y \geq 30$ için % 5.84 ve n_x, $n_y < 30$ için de % 7.16 olarak bulunmuştur. Görüleceği gibi iki örnek ortalaması farkı için de I. tip hata n_x, $n_y < 30$ için biraz büyük olmasına karşın anlamlı bir büyüklük değildir ($P>0.05$).

Anormal dağılımda σ yerine, örneklerden hesaplanan S_x ve S_y konduğu zaman, n_x, $n_y \geq 30$ olması halinde yapılan hata oranı % 5.84, n_x, $n_y < 30$ olması halinde ise yapılan hata oranı % 7.84 bulunmuştur. Bu oranlar Kan'ın (1988)'de tek ortalama için yaptığı benzer çalışmasında anormal dağılımda σ yerine S kullanıldığında bulduğu, $n > 30$ için yapılan hata oranı % 7.10 ve $n \leq 30$ için ise yapılan hata oranı % 8.20 ile paralellik göstermektedir[8]. Her iki çalışmada da örnek büyüklüklerine göre oranlar karşılaştırıldıklarında aralarındaki farkın anlamsız olduğu görülmüştür ($P > 0.05$).

Toplumun normal olması, hipotez testlerinde I. tip hatayı önemli derecede azaltmaktadır. z ve t'ye göre H_0 red oranları bu yönüyle ele alındıklarında, toplam 10000 örnek için variyans belli olmadığı zaman z red oranları % 6.59-6.95 ve t red oranları da % 5.23-5.42 gibi farklılıklar göstermektedir. Aynı oranlar σ bilindiğinde, z için % 5.04-5.89 ve t için % 5.89-6.20'dir. Benzer değerlendirme σ yerine S konularak test yapıldığı zaman da, elde edilen oranlar z testi için % 6.58-6.95 ve t için de % 5.04-5.21 gibi farklılıklar göstermektedir. Görüleceği gibi her iki testte de normal toplumdaki red oranları (I. tip hata) anormal toplumdakinden daha düşüktür ($P < 0.001$). Bu sonuçlar beklenen değerler olup, hipotez testlerinin geçerliliği için toplumun normal dağılma koşulunun gerekliliğini göstermiş olmaktadırlar[1-6,9-11,17,18].

Araş. Gör. Bülent EDİZ
Uludağ Üniversitesi Tıp Fakültesi
Biyoistatistik Bilim Dalı
16059 Görükle / BURSA

Kaynaklar

1. Kan İ, Gülesen Ö: Biyoistatistik, Yayın No: 69, Ankara, UÜ Yayınları, 1982, s. 83-222.
2. İkiz F, Püskülcü H: İstatistiğe Giriş, Yayın No: 1, İzmir, Ege Ü Yayınları, 1986, s. 115-193.
3. Çömlekçi N: İstatistik, Eskişehir: Bilim ve Tek. Kitabevi, 1982, s. 157-248.
4. Özdamar K: Biyoistatistik, Eskişehir: Bilim ve Tek. Kitabevi, 1982, s. 258-295.
5. Hays W L: Statistics, Third Edition, New York: Holt Rinehart and Winston Inc., 1981, pp. 217-294.
6. Ünver Ö, Gamgam H: Uygulamalı istatistik yöntemler, Ankara: 1986, s. 7-90.
7. Mode E B: Elements of statistics, Third Edition, New Jersey: Prentice-Hall, Inc., 1961, pp. 119-165.
8. Kan İ: Örnek büyüklüğüne bağlı olarak z ve t değerleri arasındaki ilişki, U.Ü. Tıp Fak. Dergisi, Sayı: 3, s. 1-17, 1988.
9. Serper Ö: Uygulamalı İstatistik, Cilt 2, İstanbul: Filiz Kit., 1986, s. 116-157.
10. Lapin L L: Statistics meaning and method, New York: Harcourt Brace Jovanovich Inc., 1975, pp. 203-233.
11. Çömlekçi N, Yüzer AF, Ağaoğlu E: İstatistik, Şeniş F (Editör), Ankara: Yayın No: 202, Fasikül: 3, A.Ü. Açık Öğr. Yay., 1985, s. 172-229.
12. Bury KV: Statistical models in applied science, New York: 1975, pp. 243-247.
13. Spiegel RM: Theory and problems of statistics, New York: McGraw-Hill Book Company, 1961, pp. 141-190.
14. Hoel PG: Elementary statistics, Fourth Edition, New York: John Wiley and Sons, 1976, pp. 113-194.
15. Mood AM, Graybill FA: Introduction to statistical analysis, Second Edition, New York: McGraw-Hill Book Company, 1963, pp. 123-126.
16. Dubois EN: Essential methods in business statistics, New York: McGraw-Hill Book Co., 1964, pp. 116-125.
17. Hamburg M: Basic statistics, Second Edition, New York: Harcourt Brace Jovanovich Inc., 1979, pp. 198-214.
18. Oruç M: İstatistik yöntemler, Ankara: Bizim Büro Yayınları, 1977, s. 188-239.