Eskişehir Anadolu Üniversitesi
İktisadi ve İdari Bilimler Fak. Dergisi
C: XI, S: 1-2, 1993

İKİ ORTALAMANIN TESTİNDE SATTERTHWAITE'IN YAKLAŞIK F TESTİ İLE t TESTİNİN KARŞILAŞTIRILMASI

Öğ. Gr. Dr. Embiya AĞAOĞLU*

Dağılımı normal olan bağımsız iki ana kütleden çekilen örneklemlere dayanılarak anakütle ortalamaları test edilirken, önce varyansların homojen olup olmadığı sınanmaktadır. Varyansların homojen olduğu durumda t testi, homojen olmadığı durumda Smith/Welch/Satterthwaite olarak da bilinen test uygulanmaktadır. Ancak, bu süreçte varyansların testi uygun mudur? İki ortalamanın karşılaştırılmasında birbirinden farklı iki ayrı formül kullanılmalı mı sorularına yanıt araştırılacaktır. Bu sorulara yanıt araştırılırken, testin gücü ve testin büyüklüğü (the size of the test) esas alınarak, bu iki ölçütün belirlenmesinde varyans testi de göz önüne alınacaktır.

I. GİRİŞ

Dağılımı normal olan bağımsız iki ana kütleden çekilen örneklemlere dayanılarak anakütle ortalamalarının eşitliğinin sınanması, ana kütle varyanslarının eşit olduğu varsayımı altında t testiyle yapılmaktadır. Varyanslar eşit değilse, o zaman Smith (1936), Welch (1937) ve daha sonra Satterthwaite (1946) tarafından önerilen yöntem genellikle kullanılmaktadır. Temel istatistik kitapla-

* Anadolu Üniversitesi Fen Fakültesi İstatistik Bölümü Öğretim Elemanı.

rında, bilgisayar paket programlarında ve istatistik öğretiminde de bu süreç izlenmektedir.

Örnek ortalamalarının karşılaştırılmasında, varyansların homojenliği genellikle F testiyle sınanmaktadır. Bu çalışmada, bu test ön varyans testi olarak ifade edilmektedir. Ön-varyans testiyle ilgili araştırmalar arasında Bozovich, Bancraft ve Hartley (1956)'ın araştırmaları vardır. Onlar araştırmalarında özellikle birleştirilmiş varyans üzerinde durmaktadır. Gurland ve McCullough (1962) ise, McCullough, Gurland ve Rosenberg (1960)'in geliştirdiği iki ortalamanın karşılaştırılmasında ön-varyans testinin, testin gücüne etkisini araştırmaktadır. Bu araştırmaların herbirinde Satterthwaite testi incelenmemiştir. Cochran (1951), Hudson ve Krutchkoff (1968), Davenport ve Webster (1973), Lorenzen (1987) ve Best Raynor (1987)' de Satterthwaite'ın yaklaşık F testini geniş olarak incelemektedir. Bu yazarlar, ön-varyans testinde testin anlam seviyesi $\alpha = 1$ için herhangi bir test yapmaksızın, yani varyanslar homojen kabul edilerek Satterthwaite'ın "An Always Satterthwaite test (AS)"nı incelemişlerdir. Fakat, yazarlar $H_0 : \sigma_1^2 = \sigma_2^2$ ile ilgili sorunları incelememektedir (1, 2, 3).

Ön-varyans testi, t testi ile Satterthwaite testinden hangisinin kullanılacağını çözümlemesine karşın, bazı önemli sorular ortaya çıkmaktadır. Birincisi, mevcut uygulamada iki ortalama test edilirken, varyansların homojenliğinin test edilmesi uygun mudur? İkincisi, iki örnek ortalamasının karşılaştırılmasında birbirinden farklı iki ayrı formül kullanılmalı mı? Mevcut bilgisayar paket programlarında durum nedir? sorularına yanıt araştırılacaktır.

II. TESTLERLE İLGİLİ AÇIKLAMALAR

Dağılımı normal olan iki ana kütleden çekilen bağımsız tesadüfi örnekler $x_{11}, x_{12}, \dots, x_{1n_1}$ ve $x_{21}, x_{22}, \dots, x_{2n_2}$ olsun Yani, $x_{1i} \sim N(\mu_1, \sigma_1 2)$ ve $x_{2j} \sim N(\mu_2, \sigma_2^2)$; $i = 1,2,..,n_1$ ve $j = 1,1,\dots,n_2$ dir.

İki ana kütlenin varyanslarının eşitliğinin sınanması için;

$H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 \neq \sigma_2^2$

hipotezleri kurulur ve H_0 hipotezi

$$F = s_2^2 / s_1^2 \qquad (1)$$

ile sınanır. Burada, s_1^2 ve s_2^2 örnek varyanslarıdır. α, önerilen ön-varyans testi anlam seviyesi olmak üzere, $F' > F^{\alpha/2}_{n_2-1,n_1-1}$ veya $F' < F^{1-\alpha/2}_{n_2-1,n_1-1}$ olursa, sıfır hipotezi red edilir. Burada

$$s_1^2=\sum_{i=1}^{n_1}(x_{1i}-x_1)^2/(n_1-1)\ ,\ s_2^2=\sum_{i=1}^{n_2}(x_{2i}-x_2)^2/(n_2-1)\ ,\ x_1=\sum_{i=1}^{n_1}x_{1i}/n_1\ \ \ =\sum x_{2j}/n_2 \text{ dir.}$$

Ön-varyans testi sonucu,

A. $H_0 : \sigma_1^2 = \sigma_2^2$ hipotezi kabul edilirse, o zaman ortalamalarla ilgili hipotezler ve sınama şöyledir;

Hipotezler:

$H_0^* : \mu_1 = \mu_2$ (2)

$H_1^* : \mu_1 \neq \mu_2$ veya $H_1^* : \mu_1 > \mu_2$ $H_1^* : \mu_1 < \mu_2$

Sıfır hipotezi aşağıdaki t istatistiği ile sınanır;

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\left[(n_1-1)s_1^2+(n_2-1)s_2^2\right]/\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \tag{3}$$

t istatistiğinde, δ önerilen ortalamalarla ilgili anlam seviyesi olmak üzere,

$t > t_{\delta,n_1+n_2-2}$ veya $t^2 > F^{\delta}_{1,n_1+n_2-2}$ olursa, H_o^* hipotezi red edilir.

B. $H_0 : \sigma_1^2 = \sigma_2^2$ hipotezi red edilirse, o zaman (3) nolu test istatistiği yerine Satterthwaite'ın önerdiği aşağıdaki test istatistiği (4) ile (2) nolu hipotezler test edilir (1,2);

$$t^*=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

δ, önerilen ortalamalarla ilgili aynı anlam seviyesi olmak üzere,

$t^* > t_{\delta,v}$ veya $t^{*2} > F^{\delta}_{1,v}$ olursa, H_o^* hipotezi red edilir. Burada serbestlik derecesi,

$$v = \frac{\left(\frac{1}{n_1} + \frac{u}{n_2}\right)^2}{\frac{1}{n_1^2(n_1 - 1)} + \frac{u^2}{n_2^2(n_2 - 1)}} \quad ; \quad u = \frac{s_2^2}{s_1^2} \text{ dir.}$$

III. TESTİN GÜCÜ, α ve β HATALARI

Bir hipotez testi, basit olarak hipotezin kabul ya da red edilmesine ilişkin bir karardır. İstatistik hipotezlerinde, belli bir konuda karar almak için önce sıfır hipotezi kurulur ve uygun bir test uygulanır. Sonuçta bu hipotez ya kabul edilir ya da red edilir. Sıfır hipotezi doğru olduğu halde test sonunda red edilirse, 1. tip hata (α hatası) işlenir. Sıfır hipotezi yanlış olduğu halde kabul edilirse, 2. tip hata (β hatası) işlenir. α tipi hatanan azalması, β tipi hatanın artmasına ve aksine α tipi hatanın artması, β hatasının azalmasına neden olur. Hipotez testlerinin amaçlarından biri α ve β hatalarının her ikisinin küçük olduğu testler düzenlemektedir (4,5).

Bir testin gücü basit olarak gerçekte sıfır hipotezi yanlışken onu red etme olasılığı olarak tanımlanır. Testin gücü, $1 - \beta$ olarak tanımlanır ve bir testte testin gücünün büyük olması istenir. O halde, testin gücünün büyük olması için, β hatasının küçük olması gerekir (5, 8).

İki ortalamanın eşitliğinin sınanmasında, testin gücü temel istatistik kitaplarında açıklandığı şekilde hesaplanabilir. Bu tür testin gücü hesabında sadece t sınaması göz önüne alınarak yapılmaktadır. Ancak t sınamasından önce varyansların homojenliği de test edilmektedir. Bu testin t testi gücü üzerindeki etkisi göz önüne alınmamaktadır. Bu çalışmada, ön-varyans testinin etkisi olabileceği varsayılarak ve ön-varyans testinde değişik anlam seviyeleri kullanılarak, t testi ve Satterthwaite testinin büyüklüğü ve gücünün nasıl etkilendiği araştırılacaktır (1)*

---

* *Testin büyüklüğü (size), $\lambda=0$ iken, H_0* hipotezinin red edilme olasılığıdır.
Testin gücü, $\lambda>0$ için H_0'ın red edilme olasılığına karşı gelmektedir.
*İki ortalamanın çift taraflı hipoteze göre sınınmasında, testin büyüklüğü ve testin gücü aşağıdaki formülle hesaplanır:

$$P(\text{red } H_o^*) = P\left(t^2 > F^{\delta}_{1,n_1+n_2-2} \text{ ve } F^{1-\frac{\alpha}{2}}_{n_2-1,n_1-1} < u < F^{\frac{\alpha}{2}}_{n_2-1,n_1-1}\right)$$

$$+ P\left(t^{*2} > F^{\delta}_{1,v} \text{ ve } u < F^{1-\frac{\alpha}{2}}_{n_2-1,n_1-1}\right)$$

$$+ P\left(t^{*2} > F^{\delta}_{1,v} \text{ ve } u > F^{\frac{\alpha}{2}}_{n_2-1,n_1-1}\right).$$

IV. ÖN - VARYANS TESTİ ANLAMLILIK SEVİYESİNİN DEĞİŞTİRİLMESİ

Ana kütlelerin varyanslarının eşitliği F testi ile sınanmakta ve anlam seviyesi $\alpha = 0$ ile $\alpha = 1$ arasında olabilmektedir. Şimdi α anlam seviyesine 0 - 1 arasında değerler verilerek anakütle ortalamaları arasındaki fark test edilirken, bu testleri tanımlayalım :

Önerilen ön-varyans testi anlam seviyesi $\alpha = 0$ olursa, o zaman $H_0 : \sigma_1^2 = \sigma_2^2$ hipotezi her zaman kabul edilir ve $H_0^* : \mu_1 = \mu_2$'yi test etmek için her zaman t testi yapılabilir. Aynı şekilde, $\alpha = 1$ alınırsa, her zaman $H_0 : \sigma_1^2 = \sigma_2^2$ red edilir ve $H_0^* : \mu_1 = \mu_2$ yi test etmek için Satterthwaite testi yapılabilir. α yı sıfır veya bir olarak belirlemek, $H_0 : \sigma_1^2 = \sigma_2^2$ için ön - varyans testi yapmaksızın t testi veya Satterthwaite testi yapmaya eşdeğerdir. α anlam seviyesi $0 < \alpha < 1$ aralığında iken H_0 için ön-varyans testinden sonra yapılan t testi veya Satterthwaite'ın testi "Sometimes Satterthwaite testi (SS testi)" diye ifade edilecektir. H_0 için (yani $\alpha = 0$), ön-varyans testi yapılmaksızın H_0^* için t testi yapmayı "Always t testi (AT testi)" ve H_0 için (yani ($\alpha = 1$) ön - varyans testi yapılmaksızın H_0^* için Satterthwaite testi kullanmayı Always Satterthwaite testi (AS testi) olarak ifade edilecektir (1,2).

V. AT, SS ve AS TESTLERİNİN BÜYÜKLÜKLERİ

AT, SS ve AS testlerinin büyüklükleri n_1, n_2, α, δ ve $\theta = \sigma_2^2/\sigma_1^2$ nın fonksiyonlarıdır. Burada ön - varyans testinin anlam seviyesi olan α ya değişik değerler verilerek, söz konusu testlerin büyüklüğü hesaplanarak etkisi sayısal olarak gösterilecektir. Bu amaçla

$\alpha = 0, 0.05, 0.25, 1$ olarak alınmıştır. Tüm testlerde $\delta = 0.05$ kullanılmıştır. Örnek hacimleri $n_1 - 1 = 5, 10, 50$ $n_2 - 1 = 5, 10, 50$ olarak alınarak, Ek Tablo 1'de görüldüğü gibi değişik kombinezonlarda kullanılmıştır. $\theta = 1, 2, \ldots, 10$'dur.

Ek Tablo 1'deki sayısal verilerin yorumları şöyledir; Eğer iki örnek hacmi birbirine eşitse ($n_1 = n_2$), θ'nın tüm değerleri için AT, SS ve AS testlerinin hepsinin büyüklüğü testin anlam seviyesi olan $\delta = 0.05$'e oldukça yakındır.

Örneğin, $n_1\text{-}1=5$, $n_2\text{-}1=5$; $n_1\text{-}1=10$, $n_2\text{-}1=10$; $n_1\text{-}1=50$, $n_2\text{-}1=50$ hacimli örneklerde testin büyüklükleri 0.05'e çok yakındır. Diğer taraftan, $n_1 \neq n_2$ durumunda da AS testinin büyüklükleri $\delta = 0.05$'e çok yakın değerler almaktadır.

Örnek hacimlerinin eşit olmadığı durumları ise, örneklerin varyansları da göz önüne alınarak iki grupta incelenecektir. Birincisinde, $n_1 > n_2$ ve $s_1^2 < s^2$ ise testlerin büyüklüğü, θ'nın ve ön-varyans testi α'nın farklı değerlerine göre şöyledir; $n_1\text{-}1 = 5$ ve $n_2\text{-}1 = 10$ olduğunda ortalamalar arasındaki farkın testinde SS testi, $\theta=1$ değeri hariç almak üzere, $\alpha=0.05$ ve $\theta=4$ için en düşük büyüklük (0.0332) ve $\alpha=0.05$ ve $\theta=3$ için en düşük büyüklük (0.0439) olmaktadır. Yani, SS testinin büyüklüğü θ'nın en küçük değeri ile en büyük değeri arasında minimum bir seviyeye inmektedir. AT testindeki büyüklükleri ise, θ en küçük değerinden en büyük değere giderken testin büyüklüğü $\delta=0.05$'ten uzaklaşmaktadır. AS testinde testin büyüklüğü $\delta=0.05$'e yakın olmaktadır. Örnek hacimleri $n_1\text{-}1=5$ ve $n_2\text{-}1=50$, $n_1\text{-}1=10$ ve $n_2\text{-}1=50$ olan testlerin büyüklükleri de aynı durumdadır.

Örnek hacimleri $n_1 < n_2$ ve varyansları $s_1^2 > s_2^2$ olan ortalamaların testleri incelendiği zaman şu sonuçlar gözlenmektedir. SS testlerinde testin büyüklüğü, θ'nun 2, 3 ve 4 değerlerinde testin anlam seviyesi olan $\delta=0.05$'ten uzaklaşarak en büyük değerlere örneğin 0.0818, 0.0630, 0.1182, 0.0830, 0.0848 ve 0.0640 değerlerine ulaşmaktadır. AT testinde de θ değeri arttıkça ortalama farkıyla ilgili testin büyüklüğü $\delta=0.05$'ten daha büyük değerlere doğru sürekli bir artış göstermektedir. AS testindeki durum ise şöyledir; θ'nın tüm değerleri için testin büyüklüğü $\delta=0.05$ civarında bulunmaktadır.

VI. AT, SS ve AS TESTLERİNİN GÜÇLERİ

AT, SS ve AS testlerinin gücü n_1, n_2 θ, δ, α ve $\lambda = (\mu_1-\mu_2)/[2(\sigma_1^2/n_1+\sigma_2^2/n_2)]$ nın fonksiyonlarıdır. Ön - varyans testinin anlam seviyesi α ya $\alpha=0.05$, 0.25 ve 1 değerleri verilerek testlerin güçleri hesaplanacak ve etkisi araştırılacaktır. Bunun için, $\theta=1$ ve 10 değerleri alınarak, $\lambda=0,1,\ldots,10$ ve n_1-$1=5$, n_1-$1=50$, n_2-$1=5$ ve n_2-$1=50$ örnek hacimleri kullanılacaktır. Sonuçlar Ek Tablo 2'de gösterilmiştir.

Ek Tablo 2'den görüldüğü gibi AT, SS ve AT testlerinin güçleri örnek hacimleri eşit olduğu zaman aynı θ ve λ değerleri için hemen hemen aynı sonuçlar elde edilmektedir. Örneğin, $\theta=1$, n_1-$1=5$ ve n_2-$1=5$, $\lambda=6$ için AT, SS ve AS testlerinin güçleri sırasıyla 0.5991, 0.5953, 0.5869 ve 0.5810'dur. Hatta, örnek hacimleri büyüdükçe hiç fark kalmamaktadır. Örneğin $\theta=1$, n_1-$1=50$ ve n_2-$1=50$, $\lambda=6$ için AT, SS ve AS testlerinin güçleri 0.6005'tir.

Örnek hacimleri eşit olmadığı zaman, yani $n_1 < n_2$ ve $\theta=1$ için, $\lambda=0$ hariç, AT testinin gücü diğerlerine göre biraz daha fazladır. n_1-$1=5$ ve n_2-$1=50$ de, küçük örneğin varyansı büyük örneğin varyansından daha küçüktür ve $\theta=10$ için AS testinin gücü en büyük, bunu ise sırasıyla SS ve AS tesetleri izlemektedir. n_1-$1=50$, n_2-$1=5$'te büyük örneğin varyansı diğerine göre daha küçüktür ve AT testinin gücü diğerlerine göre büyüktür. SS testleri ile AS testinin güçleri ise birbirine oldukça yakındır. Burada AT testinin $\lambda=0$ iken testin büyüklüğü 0.3801'dir. Bu değer testin anlam seviyesi $\delta=0.05$'ten oldukça büyüktür. Diğer taraftan, AS testinin $\lambda=0$ iken testin büyüklüğü 0.0512'dir ki, testin anlam seviyesi $\delta=0.05$'e çok yakındır.

Ekteki Tablo 1 ve Tablo 2'deki sayısal veriler ayrı ayrı yukarıda açıklanmıştır. Şimdi her iki tablo göz önüne alınarak özet bir açıklaması aşağıdaki gibi yapılabilir (2):

1. Örnek hacimleri eşitse, tüm n_1, n_2, $\lambda > 0$ ve $1 \leq \theta \leq 10$ için AS, SS ve AT testlerinde hemen hemen testin büyüklükleri ve testin güçleri aynıdır.

2. Örnek hacimleri eşit değil, fakat varyans oranı, θ, bire yakınsa ,o zaman AT testi δ seviyesine yakın bir testin büyüklüğünü korurken en büyük testin gücüne ulaşmaktadır.

3. Örnek hacimleri eşit değil, θ bire yakın değilse ve örnek hacmi büyük olanın varyansı daha küçükse, o zaman AT testi hâlâ tüm $\lambda > 0$ değerleri için en büyük testin gücüne sahiptir ve onu SS ve AS testleri izlemektedir. Mamafih, AT testi büyük bir testin büyüklüğü pahasına bu büyük testin gücüne ulaşmaktadır. θ'nın bazı değerleri için, SS testi testin büyüklüğünün büyük olması pahasına daha büyük testin gücüne ulaşacaktır. Mamafih, AS testi tüm $1 \leq \theta \leq 10$ değerleri için δ'ye yakın bir testin büyüklüğünü korumaktadır.

4. Örnek hacimleri farklı, θ bire yakın değil ve örnek hacmi küçük olan örneğin varyansı daha küçükse, o zaman AS testi en büyük testin gücüne sahiptir ve onu SS ve AT testleri izlemektedir. AS testi tüm $1 \leq \theta \leq 10$ değerleri için uygun bir testin büyüklüğünü korumaktadır.

VII. PAKET PROGRAMLARDA İKİ ÖRNEK ORTALAMASI

İki ortalamanın karşılaştırılmasını bilgisayar paket programları çıktıları üzerinde açıklayalım. SAS, MINITAB, SYSTAT program çıktıları aşağıdadır :

— SAS Paket Program Çıktısı —

INSTRUCT	N	MEAN	STD DEV	STD ERROR
PR 110	45.563636		10.248765	0.977181
TA	89	44.932584	9.762867	1.034861

VARIANCES	T	DF	PROB>ITI
UNEQUAL	0.4434	191.8	0.6580
EQUAL	0.4411	197.0	0.6596

FOR HO: VARIANCES ARE EQUAL, F' = 1.10 WITH 109 AND 88 DF
PROB>F' = 0.6383

— MINITAB Paket Program Çıktısı —

1. TWO SAMPLE T FOR C3

C4	N	MEAN	ST DEV	SE MEAN
1	12	35.83	7.76	2.2
2	20	43.00	17.60	3.9
95	PCT	C1	FOR	MU1-MU2 : (-18.2, 3.8)

T TEST MU1=MU 2(VS NE) : T= -1.33 P=0.19 DF=30
POOLED ST DEV=14.7

2. TWO SAMPLE T FOR C3

C4	N	MEAN	ST DEV	SE MEAN
1	12	35.83	7.76	2.2
2	20	43.00	17.60	3.9
95	PCT	C1	FOR	MU1-MU2 : (-16.4, 2.1)

T TEST MU1=MU2 (VS NE) : T= -1.59 P=0.12 DF=28

— SYSTAT Paket Program Çıktısı —

G=1.000	
N OF CASES	12
MEAN	35.833
STANDARD DEV	7.756
STD. ERROR	2.239
G=2.000	
N OF CASES	20
MEAN	43.000
STANDARD DEV	17.568
STD. ERROR	3.928

SUMMARY STATISTICS FOR X
BARTLETT TEST FOR HOMOGENEITY OF GROUP VARIANCES
CHI-SQUARE=7.227 DF=1 PROB.=0.007
OVERALL MEAN=40.313 STANDARD DEV.=14.931
POOLED WITHIN GROUPS STANDARD DEV.=14.749
T STATISTICS=1.331 PROB.=0.193

Örnek olarak alınan bu üç paket programında iki örnek ortalaması şu şekilde karşılaştırılabilir; SAS paketinde t testi yapılırken varyansların homojenliği, yani ön-varyans testi için ayrı bir komut yoktur, paket doğrudan ön-varyans testi için F değerini vermektedir .Ayrıca, varyansların homojen veya homojen olmadığı durumdaki her iki t değerleri sonuçlarını da vermektedir. Araştırıcı ön-varyans testi için benimsediği anlam seviyesini göz önüne alarak t testinin iki sonucundan birini tercih etmektedir (5).

Minitab paket program çıktısı varyansların homojen olup olmamasına göre iki ayrı tablo şeklinde t değerleriyle ilgili sonuçları verir. Varyansların homojenliğinin testi için gerekli bilgiler her iki tabloda bulunmasına karşın F değerleriyle ilgili hiç bir bilgi bulunmamaktadır. Bu nedenle varyansların testi kullanıcıya bırakılmıştır. Bu durum göz önüne alındığında, kullanıcılar her iki t değeri sonuçlarını almakta ve daha sonra, F değerinin sonucuna göre, uygun t değerini seçmektedir (7).

Systat paket program çıktısında, iki örnek ortalamasıyla ilgili gerekli aritmetik ortalama, standart sapma ve standart hata sonuçları verildikten sonra, t değeriyle ilgili tablo verilmektedir. Bu hesaplamada varyansların homojenliği Bartlett testiyle sınanarak sınama sonucuna göre uygun t formülü kullanılmaktadır (6).

VIII. SONUÇ VE ÖNERİLER

İki ortalamanın karşılaştırılmasında yapılan ön-varyans testinin uygun olmadığı sonucu, aşağıdaki açıklamalarla benimsenebilir.

Örnek hacimleri farklı ve varyans oranı bire yakınsa, t testi, anlam seviyesi δ ye yakın bir testin büyüklüğünü korurken en büyük testin gücünü sağlamaktadır. Bundan dolayı, örnek hacimlerinin farklı ve varyans oranının bire yakın olduğu biliniyorsa, o zaman t testi uygundur.

Örnek hacimlerinin eşit olduğu zaman Satterthwaite testi, SS ve t testlerinin hepsi aynı testin büyüklüğü ve testin gücüne sahiptir. Bundan dolayı, ön-varyans testi gereksizdir ve fazladan yapılmış bir işlem olmaktadır. Böylece, eşit örnek hacimlerinde, ya t testi ya da Satterthwaite testi uygundur.

Yukarıdaki açıklamalardan sonra, örnek hacimlerinin farklı olduğu ve varyans oranının bilinmediği veya birden farklı olduğunun bilindiği durumlar kalmaktadır. Örnek hacimlerinin farklı olmasında, Satterthwaite testi δ ye yakın uygun bir testin büyüklüğünü muhafaza ederken testin gücü büyüktür. Mamafih varyans oranı birden farklıyken hem SS hem de t testinin büyüklüğü yeterlidir. Bundan dolayı, Satterthwaite testi burada uygundur.

Ön-varyans testli SS testinin asla önerilmediğine dikkat ediniz. Varyans oranının tüm değerleri için SS testi iyi sonuç verirken hem t testi hem de Satterthwaite testinin kötü sonuç verdiği hiç bir örnek bileşimi yoktur. Bu, SS testinin büyüklüğünün (gücünün) t testi ve Satterthwaite testinin testin büyüklüğünün (gücünün) tartılı bir ortalaması olmasının sonucudur. Bundan dolayı, ön-varyans testi bu sorun için asla uygun değildir.

Nihayet, varyansların homojenliği testi üzerindeki ısrarlı duruş haklılık kazanmış mıdır? Bu soruya 'evet' demek mümkün görülmemektedir. Çünkü varyans oranı bilindiği zaman o sadece uygun test yönteminin seçimini etkiler. Mamafih, uygulamada, varyans oranı nadiren bilinir. Yukarıdaki tartışmalardan, Satterthwaite testi varyans oranının bilinmediği tüm durumlarda uygundur sonucu benimsenmektedir.

EK : TABLO I

İKİ ÖRNEK ORTALAMASININ KARŞILAŞTIRILMASINDA ANLAM SEVİYESİ $\delta = 0.05$ İÇİN AT, SS VE AS TESTLERİNİN BÜYÜKLÜKLERİ ($\lambda = 0$ İÇİN)

			AT test	SS test	SS test	AS test
n_1 -1	n_2 -1	θ	$\alpha = 0$	$\alpha = .05$	$\alpha = .25$	$\alpha = 1$
5	5	1	.0500	.0492	.0473	.0459
		2	.0521	.0509	.0486	.0472
		3	.0549	.0529	.0501	.0487
		4	.0574	.0544	.0511	.0499
		5	.0594	.0554	.0518	.0507
		6	.0610	.0561	.0522	.0512
		7	.0623	.0565	.0525	.0516
		8	.0635	.0567	.0527	.0519
		9	.0644	.0567	.0528	.0521
		10	.0653	.0567	.0529	.0522
5	10	1	.0500	.0500	.0507	.0507
		2	.0329	.0370	.0448	.0481
		3	.0267	.0337	.0439	.0479
		4	.0236	.0332	.0443	.0481
		5	.0218	.0339	.0452	.0484
		6	.0206	.0350	.0460	.0486
		7	.0198	.0362	.0468	.0489
		8	.0192	.0375	.0474	.0491
		9	.0187	.0387	.0479	.0493
		10	.0183	.0398	.0483	.0495
5	50	1	.0500	.0557	.0557	.0557
		2	.0110	.0308	.0516	.0550
		3	.0035	.0333	.0520	.0543
		4	.0014	.0380	.0524	.0536
		5	.0007	.0417	.0523	.0529
		6	.0004	.0443	.0520	.0523
		7	.0002	.0460	.0516	.0518
		8	.0001	.0472	.0513	.0514
		9	.0001	.0480	.0510	.0510
		10	.0001	.0485	.0508	.0508

			AT test	SS test	SS test	AS test
n_1 -1	n_2 -1	θ	$\alpha = 0$	$\alpha = .05$	$\alpha = .25$	$\alpha = 1$
10	5	1	.0500	.0500	.0507	.0507
		2	.0763	.0703	.0607	.0523
		3	.0945	.0792	.0630	.0532
		4	.1076	.0818	.0626	.0536
		5	.1174	.0813	.0615	.0537
		6	.1251	.0793	.0601	.0536
		7	.1312	.0769	.0589	.0535
		8	.1362	.0743	.0579	.0534
		9	.1403	.0719	.0570	.0533
		10	.1439	.0697	.0563	.0532
10	10	1	.0500	.0488	.0488	.0488
		2	.0513	.0504	.0496	.0493
		3	.0529	.0510	.0501	.0498
		4	.0543	.0514	.0503	.0502
		5	.0554	.0514	.0505	.0504
		6	.0562	.0514	.0506	.0505
		7	.0570	.0513	.0507	.0506
		8	.0576	.0512	.0507	.0507
		9	.0581	.0511	.0507	.0507
		10	.0585	.0511	.0507	.0507
10	50	1	.0500	.0513	.0513	.0513
		2	.0154	.0336	.0476	.0509
		3	.0069	.0388	.0492	.0505
		4	.0038	.0441	.0499	.0503
		5	.0024	.0470	.0499	.0501
		6	.0017	.0485	.0499	.0499
		7	.0012	.0492	.0499	.0499
		8	.0010	.0495	.0498	.0498
		9	.0008	.0496	.0498	.0498
		10	.0007	.0497	.0498	.0498

TABLO 1'in devamı

			AT test	SS test	SS test	AS test
50	5	1	.0500	.0557	.0557	.0557
		2	.1344	.1081	.0630	.0539
		3	.1987	.1182	.0797	.0531
		4	.2460	.1118	.0728	.0525
		5	.2819	.1000	.0672	.0522
		6	.3100	.0900	.0630	.0519
		7	.3327	.0823	.0601	.0517
		8	.3513	.0761	.0579	.0515
		9	.3669	.0713	.0564	.0514
		10	.3801	.0676	.0552	.0512
50	10	1	.0500	.0513	.0513	.0513
		2	.1153	.0848	.0540	.0512
		3	.1622	.0795	.0588	.0510
		4	.1959	.0688	.0548	.0508
		5	.2209	.0614	.0528	.0507
		6	.2402	.0570	.0517	.0506
		7	.2555	.0545	.0511	.0506
		8	.2680	.0530	.0509	.0505
		9	.2873	.0521	.0507	.0505
		10	.2870	.0515	.0506	.0504
50	50	1	.0500	.0499	.0499	.0499
		2	.0503	.0500	.0500	.0500
		3	.0507	.0500	.0500	.0500
		4	.0510	.0500	.0500	.0500
		5	.0512	.0500	.0500	.0500
		6	.0514	.0500	.0500	.0500
		7	.0515	.0500	.0500	.0500
		8	.0516	.0500	.0500	.0500
		9	.0517	.0500	.0500	.0500
		10	.0518	.0500	.0500	.0500

Kaynak : Barry K. MOSER, Gary R. STEVENS, Christian L. WATTS. **"The Two - Sample T Test Versus Satterthwaite's Approximate F Test"**, Communications in Statistics-Theory Meth., 18 (11), 3963-3975, (1989).

EK: TABLO II

İKİ ÖRNEK ORTALAMASININ KARŞILAŞTIRILMASINDA AT, SS VE AS TESTLERİNİN GÜÇLERİ

				AT test	SS test	SS test	AS test
θ	n_1 -1	n_2 -1	λ	$\alpha = 0$	$\alpha = .05$	$\alpha = .25$	$\alpha = 1$
1	5	5	0	.0500	.0492	.0473	.0459
			1	.1482	.1463	.1421	.1390
			2	.2491	.2464	.2403	.2360
			3	.3474	.3441	.3368	.3315
			4	.4396	.4360	.4280	.4221
			5	.5239	.5200	.5116	.5056
			6	.5991	.5953	.5869	.5810
			7	.6653	.6615	.6534	.6476
			8	.7224	.7189	.7112	.7059
			9	.7714	.7680	.7610	.7560
			10	.8128	.8097	.8033	.7988
1	5	50	0	.0500	.0557	.0557	.0557
			1	.1710	.1709	.1551	.1449
			2	.2875	.2874	.2560	.2348
			3	.3982	.3981	.3527	.3215
			4	.5020	.4987	.4421	.4031
			5	.5936	.5874	.5526	.4784
			6	.6723	.6636	.5932	.5470
			7	.7383	.7276	.6549	.6086
			8	.7932	.7809	.7077	.6633
			9	.8379	.8246	.7527	.7116
			10	.8736	.8601	.7907	.7538
1	50	50	0	.0500	.0500	.0500	.0500
			1	.1677	.1677	.1677	.1677
			2	.2884	.2882	.2882	.2882
			3	.4035	.4033	.4033	.4033
			4	.5083	.5083	.5083	.5083
			5	.6005	.6005	.6005	.6005
			6	.6794	.6794	.6794	.6794
			7	.7455	.7455	.7455	.7455
			8	.7999	.7999	.7999	.7999
			9	.8440	.8440	.8440	.8440
			10	.8794	.8793	.8793	.8793

TABLO 2'nin devamı

θ	n_1 -1	n_2 -1	λ	AT test $\alpha = 0$	SS test $\alpha = .05$	SS test $\alpha = .25$	AS test $\alpha = 1$
10	5	5	0	.0653	.0567	.0529	.0522
			1	.1695	.1476	.1404	.1395
			2	.2717	.2379	.2287	.2277
			3	.3682	.3238	.3144	.3134
			4	.4569	.4062	.3955	.3945
			5	.5367	.4811	.4706	.4697
			6	.6075	.5490	.5391	.5384
			7	.6694	.6101	.6009	.6003
			8	.7231	.6642	.6560	.6555
			9	.7691	.7120	.7047	.7043
			10	.8082	.7538	.7475	.7471
10	5	50	0	.0001	.0485	.0508	.0508
			1	.0010	.1493	.1600	.1601
			2	.0023	.2500	.2714	.2717
			3	.0066	.3453	.3783	.3788
			4	.0137	.4320	.4766	.4774
			5	.0240	.5088	.5643	.5653
			6	.0383	.5755	.6410	.6421
			7	.0565	.6328	.7066	.7080
			8	.0789	.6817	.7620	.7636
			9	.1055	.7231	.8081	.8099
			10	.1360	.7583	.8462	.8481
10	50	5	0	.3801	.0676	.0552	.0512
			1	.5756	.1475	.1359	.1324
			2	.7107	.2276	.2176	.2149
			3	.8034	.3058	.2975	.2955
			4	.8669	.3806	.3704	.3724
			5	.9101	.4509	.4458	.4446
			6	.9395	.5160	.5122	.5113
			7	.9594	.5757	.5728	.5722
			8	.9728	.6297	.6276	.6272
			9	.9818	.6781	.6766	.6763
			10	.9878	.7214	.7202	.7200

				AT test	SS test	SS test	AS test
10	50	50	0	.0500	.0500	.0500	.0500
			1	.1701	.1662	.1662	.1662
			2	.2907	.2853	.2853	.2853
			3	.4054	.3993	.3993	.3993
			4	.5096	.5033	.5033	.5033
			5	.6011	.5950	.5949	.5949
			6	.6795	.6737	.6737	.6737
			7	.7451	.7400	.7399	.7399
			8	.7993	.7948	.7947	.7947
			9	.8432	.8394	.8393	.8393
			10	.8785	.8751	.8751	.8751

KAYNAKÇA

Barry K. MOSER, Gary R. STEVENS, Christian L. WATTS, "The Two Sample T Test Versus Satterthwaite's Approximate F Test", Communications in Statistics - Theory and Methods, 18 (11), 3963-3975, 1989.

Barry K. MOSER, Gary R. STEVENS, "Homogeneity of Variance in the Two - Sample Means Test", The American Statistician, February 1992, Vol. 46, No. 1, pp. 19-21.

J. GURLAND and R. S. Mc. CULLOUGH, "Testing Equality of Means After a Preliminary Test of Equality of Variance", Biometrika, 43,3 and 4, 43-417, 1973.

Richard P. RUNYON and Audrey HABER, Fundamentals of Behavioral Statistics Fourth edition, 1980, pp. 308-310.

Richard I. LEVIN, Statistics for Management, Fourth edition, 1987, 1987, pp. 374-376, 386-388, 424-425.

Leland WILKINSON, Systat, 1984.

Thomas A. RYAN, JR. Brian L. JOINER, Barbara F. RYAN, Minitab, 1982.

William MENDENHALL, Dennis D. WACKERLY, Richard L. SCHEAFFER, Mathematical Statistics With Applications, Fourth edition, pp. 453-459, 467-469.