

TEMEL İSTATİSTİK

Varyans Analizi I

Prof. Dr. Ezel Tavşancıl

Giriş

- Varyans analizi =ANOVA (Analysis of Variance)
- İki ya da daha çok evrene ait ortalamaların karşılaştırılması
- Bağımlı değişkene işaret eden örneklemelerin bağımsız değişken açısından karşılaştırılması üzerine kuruludur.
- Tek yönlü ANOVA: Bir çalışmada ya da deneyde bağımlı değişken üzerindeki etkisi araştırılan tek bir bağımsız değişken olduğunda yapılır (Bağımsız değişken = faktör).
- Dört farklı eğitim uygulanmış grubun sosyal beceri düzeyleri karşılaştırılmak istensin (Uygulanan eğitime göre sosyal beceri değişiyor mu?)
 - Bağımsız değişken (faktör): Eğitim Türü
 - Bağımlı değişken: Sosyal Beceri
 - Bağımsız değişkenin (faktörün) her bir durumu: Düzey (k) ya da işlem

! Bağımsız değişkenin farklı durumları, puanlarda manidar farklılık yaratıyorsa, buna faktörün işlem etkisi denir.

!! T testi ile benzer mantığa dayanır; t testinde iki örneklem ortalamaları (\bar{X}_1 ve \bar{X}_2) arasındaki farkı, iki ortalama arasındaki standart hatayla karşılaştırarak değerlendiriyorduk.

ANOVA'da ise t testinin bir uzantısı olarak iki ya da daha çok ortalama karşılaştırılır.

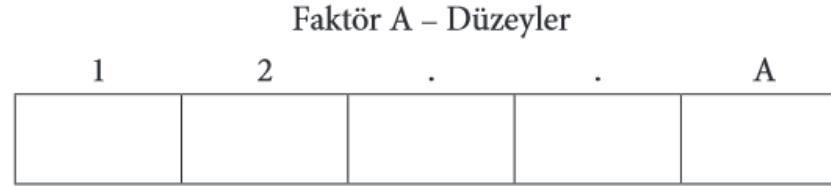
İlişkisiz iki grup ortalaması arasındaki farkın anlamlılığı için yapılacak ANOVA ile hesaplanacak F değerinin karekökü, aynı veriler için yapılacak ilişkisiz t-testi ile bulunacak t değerine eşittir ve manidarlık düzeyi (p) değişmez, dolayısıyla sonuç değişmez. İki grup için ANOVA uygulanmış ve iki ortalama puan arasındaki fark anlamlı bulunmuşsa, t-testinde olduğu gibi ortalama değerlerin büyüklüklerine bakılarak yorum yapılır.

Bu ders üzerinde çalışacağımız Tek Yönlü Varyans Analizinde

- Çoklu örneklem ortalamaları ($\bar{X}_1, \bar{X}_2, \bar{X}_3, \bar{X}_4, \dots, \bar{X}_k$) arasındaki değişkenlik, çoklu örneklem içinindeki değişimle karşılaştırır.
- Bir bağımsız değişken vardır.
- Bağımsız değişkenin (faktörün) düzeyleri, işlem gruplarını (örneklemeleri) gösterir.
- Güçlü bir parametrik test; bu nedenle normallik, varyansların homojenliği gibi varsayımların sağlanmadığı durumlarda da kullanılabilir.

Hangi Durumlarda Kullanılır?

A bağımlı değişken üzerinde etkisi gözlenen faktör olmak üzere bir yönlü ANOVA için desen, Şekil 4.2'de gösterilmiştir. Burada A'nın her bir düzeyi, bir bağımsız değişkenin farklı deneysel işlem koşullarını ya da bir denek değişkeni için oluşan doğal grupları gösterir.



Şekil 4.2. Bir Yönlü ANOVA İçin Desen

X hastası kişilerden yansız olarak seçilen üç grupta, aynı ilacın üç farklı dozda verilmesinin hastalığın iyileşmesine olan etkilerinin incelendiği çalışma bir yönlü ANOVA'yı çağırır. Burada X hastalığına ilişkin ölçülen değerler bağımlı değişkeni, ilaç dozu (üç farklı düzeye sahip) ise bağımsız değişkeni gösterir. Yansız olarak seçilen bir grup ilköğretim okulu öğrencisinin bir konudaki performansının, tutumunun, kaygısının ya da ağırlık, boy gibi özelliklerinin örneğin sosyo-ekonomik düzeye (alt-orta-üst) göre anlamlı bir şekilde farklılaşıp farklılaşmadığının incelenmesi durumunda da ANOVA kullanılabilir.

Araştırma sorusu t-testinde olduğu gibi iki türlü yazılabilir:

- Öğrencilerin sürekli kaygıları uygulanan yöntem (A-B-C) göre anlamlı bir farklılık göstermekte midir?
- Öğrencilerin sürekli kaygıları ile uygulanan yöntem arasında anlamlı bir ilişki var mıdır?

Varsayımlar

1. Bağımlı değişkenin ölçüldüğü ölçek en az eşit aralık ölçeği düzeyindedir.
2. k örneklem bağımsızdır ve evrenlerden yansız olarak seçilir.
3. Örneklemelerin seçildiği evrenlere ait puanların dağılımı normaldir.
4. k evrene ait varyanslar homojendir.
5. Bağımsız değişken kategoriktir.

Hipotezler

- Null Hipotez:

- $H_0: \mu_1 = \mu_2 = \dots = \mu_k$

Bu hipotez, k düzeyli bir faktörlü ANOVA için geçerlidir. H_0 tüm durumlardaki örneklem ortalamalarının aynı evren ortalamasını gösterdiğini açıklar ve bunun için örneklem ortalamaları eşit olmalıdır.

- Araştırma Hipotezi:

Araştırma hipotezi (H_a) ise çeşitli μ 'lerin eşit olmadığını ya da en az bir ortalamanın farklı olduğunu açıklar:

- H_1 : En az iki evren ortalaması birbirinden farklıdır YA DA bütün μ 'ler eşit değildir.

Varyansların Homojenliği Varsayımı

- Grupların varyansının birbirine eşit olması (homojenlik) şartının sağlanıp sağlanmadığını kontrol etmek için kullanılan yöntemlerden biri: F maksimum testi.
- F değeri, varyansı en büyük olan grubun varyansının, varyansı en küçük olan grubun varyansına bölünmesi ile elde edilir:
 - $F = \frac{S_{max}^2}{S_{min}^2}$
 - Elde edilen F değeri; seçilen güven düzeyi (alfa değeri) ve her bir grup için serbestlik derecesi (n_1 ve n_2) karşılığı olan tablo değeri ile karşılaştırılır (Ek F).
 - Hesaplanan değer tablo değerinden büyük değilse, varyanslar homojen yani eşittir
 - Hesaplanan değer, tablo değerinden büyükse, varyanslar homojen değil, farklı

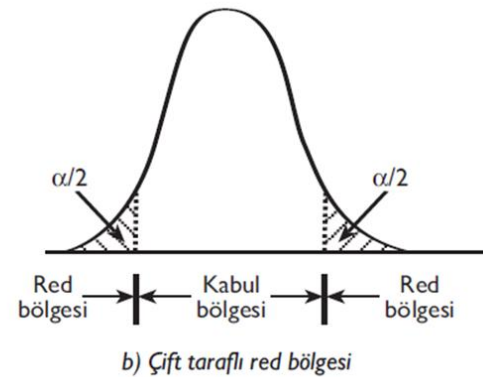
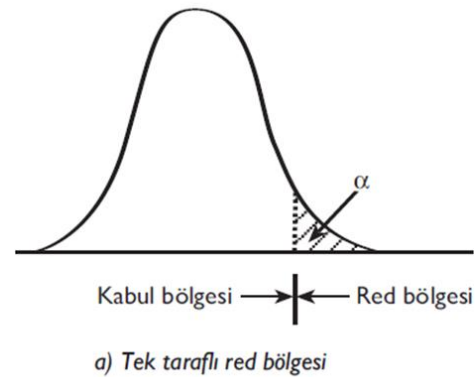
F İstatistiği ve Çoklu Karşılaştırmalar (Post Hoc)

- ANOVA'nın temeli F oranıdır.
- İki ya da daha fazla örneklem ortalamasının farklı μ 'leri temsil edip etmediğini belirlemek için tüm örneklem ortalamalarının aynı anda karşılaştırılmasını sağlar.
- Hesaplanan F değeri, belli alfa düzeyi ve serbestlik derecesine göre belirlenen kritik değer (tablo değeri) ile karşılaştırılarak karar verilir.
 - **Hesaplanan F değeri manidar değilse**, örneklem ortalamalarından **herhangi ikisi arasında manidar farklılık olmadığını** ve **tüm ortalamaların aynı μ 'yü** temsil ettiğini gösterir.
 - Hesaplanan F değeri, **manidar ise** ortalamalar arasında **en az bir farkın manidar olduğunu**, yani **en az iki ortalamanın farklı μ 'leri temsil ettiğini** gösterir. Bu durumda, bu farkın hangi ortalamalar arasında olduğunu anlamak için ikili karşılaştırmalar yapılır (çoklu karşılaştırma testleri (post hoc test)).

F İstatistiğinin Öğeleri

- Varyans analizi (ANOVA), varyansları analiz eder; varyansı kısımlara ayırarak puanlar arasındaki farkları ölçer.
- Puanlardaki toplam varyansın iki kaynağı vardır:
 - **Gruplararası (açıklanan varyans):**
 - Bağımlı değişkendeki toplam varyansın bağımsız değişken tarafından açıklanan miktarıdır.
 - Gruplararası kareler ortalamasına eşittir (KO_{GA}), ve bu değer bir faktörün düzeyleri arasında ortaya çıkan puanlardaki değişkenliğin bir tahminidir.
 - Bu amaçla faktörün her bir düzeyindeki ortalamaların, çalışmadaki tüm puanların genel ortalamasından sapması ile hesaplanır.
 - Örneklem ortalamalarının genel ortalamadan sapmalarının karesi örneklem ortalamalarının birbirinden ne kadar farklılaştığını gösterir. Böylece KO_{GA} , bir faktördeki ortalamaların birbirlerinden nasıl farklılaştığını gösterir.
 - **Gruplarıçi (açıklanamayan varyans):**
 - Faktörün her düzeyinin kendi içindeki değişkenliğini gösterir. (Farklı eğitim uygulaması gruplarındaki sosyal beceri düzeylerinin değişkenliği)
 - Gruplar arasındaki farkın bir fonksiyonu değildir.
 - Hata varyansının (error/residual variance) bir tahminidir (KO_{Gi}).
 - Farklı uygulamaların (deney durumlarının) kendi içindeki farklılıkların bir tahminidir.
 - O durumun ortalaması etrafındaki **her bir durum içindeki puanların ortalama değişkenliğidir**. Bunun için bir faktörün her bir düzeyindeki varyanslar hesaplanarak ortak varyans (pooled variance) elde edilir.
 - KO_{Gi} , her bir durumun ortalaması etrafındaki, her durumdaki puanların ortalama değişkenliğidir.

- Eğer H_0 doğru ise KO_{GA} ve $KO_{Gİ}$ 'nin eşit ya da benzer olacağı tahmin edilebilir. Bu durumda, grup ortalamaları arasındaki değişkenlik sadece örneklem hatasına bağlı olacaktır (istendiği gibi bağımlı değişken (deneysel uygulamadan ya da incelenen kategorik değişkenden) değil.).
- Eğer H_0 yanlış ise KO_{GA} 'nın, $KO_{Gİ}$ 'den büyük olacaktır. Bu durumda da bazı evren ortalamaları birbirinden farklı olacaktır.



F Dağılımı

- ANOVA'nın temel aldığı dağılımdır.
- F dağılımı, sürekli bir dağılımdır.
- Aynı normal evrenden seçilen veya aynı varyansa sahip normal evrenlerden seçildiği varsayılan iki örnekleme ait varyansların birbirine oranının ($F = \frac{S_{max}^2}{S_{min}^2}$) 1'e eşit olması beklenir. Bu hesaplanan oran her zaman 1'e eşit olmaz ve bir dağılım gösterir.
- F dağılımı, 0 ile sonsuz arasında değişim gösterir.
- Formülden görüleceği gibi (varyansların oranlanması), örneklem büyüklüğüne (serbestlik derecesine) bağlıdır.

- Varyanslar bilindiğinde, büyük olan varyans değeri küçük olana bölünür ($F = \frac{KO_{GA}}{KO_{Gİ}}$)
 - H_0 doğru ise F oranının 1'e eşit olacağı tahmin edilir ($KO_{GA} = KO_{Gİ}$)
 - H_0 yanlış ise F oranı 1'i aşacaktır ($KO_{GA} > KO_{Gİ}$)
- Her F dağılımında, pay ve paydaya karşılık gelen iki farklı serbestlik derecesi vardır. ANOVA'da,
 - KO_{GA} (pay) için serbestlik derecesi grup sayısından bir çıkarılması ($sd = k - 1$)
 - $KO_{Gİ}$ (payda) için serbestlik derecesi bütün gruptaki gözlem sayısından grup sayısının çıkarılması ($sd = N - k$)
 - Örn: Her grupta 10 kişinin yer aldığı 6 gruptan oluşan bir çalışmada serbestlik derecesi KO_{GA} için $6 - 1 = 5$ ve $KO_{Gİ}$ için $60 - 6 = 54$ 'tür.

F Oranının Hesaplanması

ANOVA'da varyans (S_X^2) kareler toplamının (KT) serbestlik derecesine (sd) bölünmesi ile elde edilir. Varyans, ANOVA'da kareler ortalaması (KO) olarak bilinir:

$$S_X^2 = \frac{\sum (X - \bar{X})^2}{N - 1} = \frac{KT}{sd} = KO;$$

Burada serbestlik derecesi N-1'dir.

Gruplararası kareler ortalaması (KO_{GA}), gruplararası kareler toplamının (KT_{GA}) gruplararası serbestlik derecesine (sd_{GA}) bölünmesi ile hesaplanır. Bunun gibi, gruplarıçi kareler ortalaması (KO_{Gi}) da gruplarıçi kareler toplamının (KT_{Gi}), gruplarıçi serbestlik derecesine (sd_{Gi}) bölünmesi ile bulunur. KO_{GA} 'nın, KO_{Gi} 'na oranlanması ile de F değeri bulunur. Aşağıda tek yönlü ANOVA için bir özet tablo verilmektedir.

Tablo 9.1: Tek Yönlü ANOVA Özet Tablosu

Varyansın Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F
Gruplararası	KT_{GA}	Sd_{GA}	KO_{GA}	F_{Hesap}
Grup içi	KT_{Gi}	Sd_{Gi}	KO_{Gi}	
Toplam	KT_{Top}	Sd_{Top}		

Tablo 4.4. İlişkisiz Ölçümler İçin Bir Yönlü ANOVA (Formüller)

Varyansın Kaynağı	Kareler Toplamı (KT)	Serbestlik Dereccesi (sd)	Kareler Ortalaması (KO)	F-Oranı
Gruplararası	KT_A	$A-1$	$[KT_A/A-1] = KO_A$	KO_A/KO_e
Gruplarıçi	KA_e	$n-A$	$[KT_e/n-A]=KO_e$	
Toplam	KT_T	$n-1$		

ÖRNEK II

(Normalde örneklem büyüklüğü parametrik test için uygun değil ama hesaplama kolaylığı açısından yapıyoruz)

Yaşın yaratıcılık üzerindeki etkisinin incelendiği bir çalışmadan elde edilen veriler tabloda verilmektedir:

Yaş 4	Yaş 7	Yaş 10	Yaş 13
3	9	9	7
5	11	12	7
7	14	9	6
4	10	8	4
3	10	9	5

- F değerini hesaplayın. ANOVA özet tablosunu hazırlayın.
- Manidarlık düzeyi .05'de, hesaplanan F değeri için karar verin.
- Yaratıcılık puanlarında açıklanabilen varyans miktarını hesaplayarak yorumlayın?

ÇÖZÜM

	Yaş 4	Yaş 7	Yaş 10	Yaş 13	
	3	9	9	7	
	5	11	12	7	
	7	14	9	6	
	4	10	8	4	
	3	10	9	5	
$\sum X$	22	54	47	29	152
$\sum X^2$	108	598	451	175	1332
n	5	5	5	5	20
\bar{X}	4.4	10.8	9.4	5.8	30.4

Kareler Toplamlarının Hesaplanması

1. Toplam kareler toplamının hesaplanması (KT_{Top}) için şu formül kullanılır:

$$KT_{Top} = \sum X_{Top}^2 - \left(\frac{(\sum X_{Top})^2}{N} \right)$$

Bu formüle tabloda hesaplanan değerler yerleştirilerek KT_{Top} hesaplanır.

$$KT_{Top} = 1332 - \frac{152^2}{20} = 1332 - 1155.2 = 176.8$$

2. Gruplararası kareler toplamının (KT_{GA}) hesaplanmasından da şu formül kullanılır:

$$KT_{GA} = \sum \left(\frac{(\text{sütündakipuanlartoplamı})^2}{\text{sütündakipuanadedi}} \right) - \left(\frac{(\sum X_{Top})^2}{N} \right)$$

Bu formüle tablodaki her sütun için hesaplanan değerler ve genel X toplamının karesi yerleştirilerek KT_{GA} hesaplanır.

$$KT_{GA} = \left(\frac{(22)^2}{5} + \frac{(54)^2}{5} + \frac{(47)^2}{5} + \frac{(29)^2}{5} \right) - \frac{152^2}{20} = 134.8$$

$$KT_{Gi} = KT_{GT} - KT_{GA} = 176.8 - 134.8 = 42.00$$

ÇÖZÜM

Kareler Ortalamasının Hesaplanması

Daha önce hesaplanan gruplararası ve gruplarıçi kareler toplamının kendi serbestlik derecelerine bölünmesi ile gruplararası kareler ortalaması (KO_{GA}) ve gruplarıçi kareler ortalaması (KO_{Gi}) elde edilir.

$$KO_{GA} = \frac{KT_{GA}}{sd_{GA}} \quad \text{Formül 9.7} \quad KO_{Gi} = \frac{KT_{Gi}}{sd_{Gi}}$$

Gruplararası ve gruplarıçi kareler toplamı ile serbestlik dereceleri formüldeki yerlerine konulduğunda, kareler ortalaması aşağıdaki gibi bulunur. Serbestlik derecesi, gruplarıçi için $N-K=20-4=16$, gruplararası için $k-1=4-1=3$ 'tür.

$$KO_{GA} = \frac{KT_{GA}}{sd_{GA}} = \frac{134.8}{3} = 44.93 \quad \text{ve} \quad KO_{Gi} = \frac{KT_{Gi}}{sd_{Gi}} = \frac{42}{16} = 2.63$$

ÇÖZÜM

F Oranının Hesaplanması

F değerinin hesaplanması için yukarıda bulunan KO_{GA} ve KO_{Gi} 'nin birbirine oranı alınır. Aşağıda F için hesaplama formülü verilmektedir:

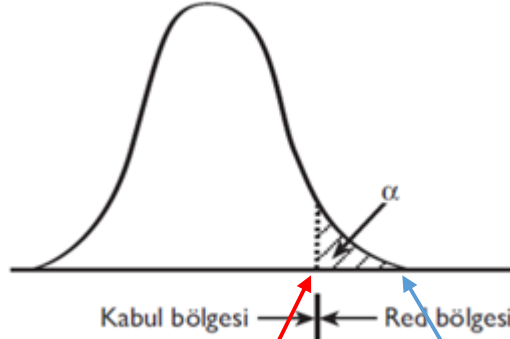
$$F = \frac{KO_{GA}}{KO_{Gi}}$$

Bu çalışmada, $KO_{Gi} = 2.63$ ve $KO_{GA} = 44.93$

Buna göre $F = \frac{44.93}{2.63} = 17.117$

Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F
Gruplararası	134.80	3	44.93	17.12
Gruplarıçi	42.00	16	2.63	
Toplam	176.80	19		

YORUMLAMA



$$F_{kritik}=3.24$$

Hesaplanan F değeri=17.12; red alanında

- $F_{(3,16)} = 17.12$ ve $F_{kritik} = 3.24$ 'tür.
- Hesaplanan F değeri, kritik F değerinden büyüktür, null hipotezi reddetme alanındadır. Bu nedenle, F değeri manidardır.
- Pratik anlamda, ortalamalar arasında fark yoktur, diyen null hipotez yanlışlanır. Karşılaştırılan grup ortalamalarından en az ikisi arasında manidar fark vardır.

Ek-F tablo devamı

14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53
	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48
	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42
	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38
	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34
	8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37
19	4.37	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31
	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28
	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25
	8.02	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23
	7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20
	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18
	7.82	5.61	4.72	4.22	3.90	3.67						
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16
	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.21	3.13	3.05	2.99

KAYNAKLAR

- Büyüköztürk, Ş. (2018). *Sosyal bilimler için veri analizi el kitabı*. Pegem Akademi: Ankara.
- Büyüköztürk, Ş., Çokluk, Ö. ve Köklü, N. (2018). *Sosyal bilimler için istatistik*. Ankara: Pegem Akademi.