IST3002 Deney Tasarımı

Bir yönlü ANOVA

Fatih Kızılaslan

Marmara Üniversitesi

2019-2020 Bahar

Bir Yönlü (One-way) ANOVA

Bir faktör ve a faktör düzeyinden oluşan ANOVA için matematiksel model

$$y_{ij} = \mu_i + \varepsilon_{ij}, \; i = 1, ..., a, \; j = 1, ..., n \tag{1}$$

biçiminde veya $\mu_i = \mu + \tau_i$, $i = 1, ..., a$ olmak üzere

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \; i = 1, ..., a, \; j = 1, ..., n \tag{2}$$

biçiminde ifade edilir. Burada,

y_{ij} : faktörün i. düzeyindeki j. gözlem (yanıt) değeri

μ_i : faktörün i. düzeyinin ortalaması

μ : tüm faktör düzeylerinin genel ortalaması

τ_i : faktörün i. düzeyinin etkisi

ε_{ij} : faktörün i. düzeyindeki j. yanıt değişkenine ilişkin rasgele hata terimi

gösterir. (1) modeli Ortalama ve (2) modeli Etki modeli olarak adlandırılır.

Modelin Varsayımları

Bir yönlü ANOVAda analizler yapabilmek için bazı varsayımlara ihtiyacımız vardır.

ε_{ij}, $i = 1, .., a$, $j = 1, ..., n$ hata terimleri birbirinden bağımsız, 0 ortalama ve sabit σ^2 varyanslı normal dağılıma sahiptir ($\varepsilon_{ij} \sim N(0, \sigma^2)$).

Bu durumda, $y_{ij} \sim N(\mu + \tau_i, \sigma^2)$ $i = 1, .., a$, $j = 1, ..., n$ olur.

(2) ile verilen model faktör düzeylerinin etkisine τ_i göre iki gruba ayrılır.

1. Sabit etkili model (Fixed effects model)
2. Rasgele etkili model (Random effects model)

İlk olarak **Sabit etkili modeli** inceleyeceğiz. Sabit etkili modelde

$$\mu = \frac{1}{a}\sum_{i=1}^{a}\mu_i \text{ olduğundan } \sum_{i=1}^{a}\tau_i = 0 \text{ olur.}$$

Hipotezler

Bir yönlü ANOVA için sıfır ve alternatif hipotezleri aşağıdaki gibi olur.

H_0 : Tüm faktör düzeylerinin ortalamaları birbirine eşittir.
H_1 : En az bir faktör düzeyinin ortalaması diğer faktör düzeylerinin ortalam

veya

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = ... = \mu_a \\ H_1 &: \mu_i \neq \mu_j \text{ en az bir } (i,j) \text{ için} \end{aligned}$$

veya

$$\begin{aligned} H_0 &: \tau_1 = \tau_2 = ... = \tau_a = 0 \\ H_1 &: \tau_i \neq 0 \text{ en az bir } i \text{ için} \end{aligned}$$

Kareler toplamının parçalanışı

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \overline{y}_{..})^2 \text{ Toplam kareler toplamı}$$

$$SS_{Deneme} = n\sum_{i=1}^{a}(\overline{y}_{i.} - \overline{y}_{..})^2 \text{ Faktörün düzeyleri (denemeler) arası kareler topla}$$

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \overline{y}_{i.})^2 \text{ Faktörün düzeyleri (denemeler) içi kareler toplamı}$$

olmak üzere

$$SS_T = SS_{Deneme} + SS_E$$

dır.

Burada, $\overline{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} y_{ij}$ faktörün i. düzeyindeki gözlemlerin ortalaması ve $\overline{y}_{..} = \frac{1}{a}\sum_{j=1}^{n} \overline{y}_{i.} = \frac{1}{a\ n}\sum_{j=1}^{n} y_{ij}$ tüm gözlemlerin ortalamasıdır.

Cochran Teoremi

Z_i, $i = 1, ..., v$ birbirinden bağımsız, standart normal dağılıma sahip rasgele değişkenler ($Z_i \sim N(0,1)$, $i = 1, ..., v$) ve Q_i, $i = 1, ..., s$, $(s \leq v)$ rasgele değişkenleri v_i serbestlik derecesine sahip olmak üzere

$$\sum_{i=1}^{v} Z_i^2 = Q_1 + Q_2 + ... + Q_s$$

olsun. Bu durumda, $Q_1, Q_2, ..., Q_s$ rasgele değişkenlerinin birbirinden bağımsız sırasıyla $v_1, v_2, ..., v_s$ serbestlik dereceli ki-kare dağılımına sahip olması için gerek ve yeterli koşul

$$v = v_1 + v_2 + ... + v_s$$

olmasıdır.

$\varepsilon_{ij} \sim N(0, \sigma^2)$ $i = 1, ..., a$, $j = 1, ..., n$ ve birbirlerinden bağımsız olması varsayımı ve H_0 hipotezi doğru olduğunda

$$\frac{SS_T}{\sigma^2} \sim \chi^2_{N-1}, \quad \frac{SS_{Deneme}}{\sigma^2} \sim \chi^2_{a-1} \text{ ve } \frac{SS_E}{\sigma^2} \sim \chi^2_{N-a}$$

olarak bulunur.

Cochran Teoreminden SS_{Deneme}/σ^2 ve SS_E/σ^2 rasgele değişkenlerinin birbirinden bağımsız olduğunu elde ederiz.

H_0 hipotezinin doğruluğu altında yani faktörün düzeyleri arasında fark yok iken

$$F_0 = \frac{SS_{Deneme}/(a-1)}{SS_E/(N-a)} = \frac{MS_{Deneme}}{MS_E} \sim F_{a-1,N-a}$$

olarak bulunur. Böylece, H_0 hipotezi için test istatistiği olarak F_0 kullanabiliriz.

Sonuç

Eğer $F_0 > F_{a-1,N-a,\alpha}$ olur ise H_0 hipotezi red edilir, yani faktör düzeyleri arasında α anlamlılık düzeyinde bir farklılık vardır.

ANOVA Tablosu

Değişim kaynağı	Kareler toplamı	Serbestlik derecesi	Kareler ortalaması	F test değeri
Denemeler	SS_{Deneme}	$a-1$	$MS_{Deneme} = \frac{SS_{Deneme}}{a-1}$	$F_0 = \frac{MS_{Dene}}{MS_E}$
Hata	SS_E	$N-a$	$MS_E = \frac{SS_E}{N-a}$	
Toplam	SS_T	$N-1$		

Burada

$$SS_{Deneme} = \frac{1}{n}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N} \text{ ve } SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N},$$

$N = a\ n$, $y_{i.} = \sum_{j=1}^{n} y_{ij}$, $y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}$.

Örnek 1

Montgomery, Design and Analysis of Experiments (9th Edition) kitabından Örnek 3.1 (sayfa 74).

Örnek 2

Dört farklı sürücü kursunun öğrencilerin ehliyet sınavındaki notlarına etkisinin önemli olup olmadığı konusunda bir araştırma yapılmak isteniyor. Rasgele olarak her bir kursa 6 şar öğrenci atanarak deneye başlanmıştır. Varyans analizi tablosonu oluşturarak sürücü kurslarının notlara olan etkisibi önemliliğini $\alpha = 0.05$ düzeyinde test ediniz.

	Kurslar			
	A	B	C	D
	70	50	80	90
	65	52	82	92
	75	60	77	82
	72	62	88	86
	74	54	83	88
	68	55	85	84
Toplam	424	333	495	522
Ortalama	70.66	55.5	82.5	87

Örnek 2 Çözüm

Hipotezlerimiz

$$
\begin{aligned}
H_0 &: \mu_1 = \mu_2 = \mu_3 = \mu_4 \iff \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0 \\
H_1 &: \mu_i \neq \mu_j \text{ en az bir } (i,j) \text{ için } \iff \tau_i \neq 0 \text{ en az bir } i \text{ için}
\end{aligned}
$$

$$
\begin{aligned}
SS_T &= \sum_{i=1}^{4}\sum_{j=1}^{6} y_{ij} - \frac{y_{..}^2}{24} = 70^2 + \ldots + 84^2 - \frac{(1774)^2}{24} = 3889.83 \\
SS_{Deneme} &= \frac{1}{6}\sum_{i=1}^{4} y_{i.}^2 - \frac{y_{..}^2}{24} = \frac{424^2 + 333^2 + 495^2 + 522^2}{6} - \frac{(1774)^2}{24} \\
&= 3567,5 \\
SS_E &= SS_T - SS_{Deneme} = 322.33
\end{aligned}
$$

ANOVA tablosu

Değişim kaynağı	Kareler toplamı	Serbestlik derecesi	Kareler ortalaması	F test değeri
Denemeler	$3567,5$	$4-1=3$	1189.167	$F_0 = 73.78$
Hata	322.33	$24-4=20$	16.1165	
Toplam	3889.83	$24-1=23$		

F tablosundan $F_{a-1,N-1,\alpha} = F_{3,20,0.05} = 3.10$ bulunur.
$F_0 = 73.78 > F_{3,20,0.05} = 3.10$ olduğundan H_0 hipotezi $\alpha = 0.05$ anlamlılık düzeyinde red edilir.

Sonuç

Böylece, "bu kurslardaki öğrencilerin sınav notlarının ortalamaları birbirinden farklıdır" veya "bu sürücü kursları arasında anlamlı bir farklılık vardır" veya "sürücü kursu öğrencilerin bu sınavdan aldığı notu etkiler".

R programında çözüm

Kodlar:

x<-
c(70,65,75,72,74,68,50,52,60,62,54,55,80,82,77,88,83,85,90,92,82,86,88,84)
faktör<-c(rep(”A”,6),rep(”B”,6),rep(”C”,6),rep(”D”,6))
veri<-data.frame(x,faktör)
veri
anova<-aov(x˜faktör,data=veri)
summary(anova)

Sonuç:

```
> anova<-aov(x~faktör,data=veri)
> summary(anova)
            Df Sum Sq Mean Sq F value   Pr(>F)    
faktör       3   3567  1189.2   73.78 5.43e-11 ***
Residuals   20    322    16.1                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```