آنالیز واریانس

گاهی اوقات پیش می آید که بخواهیم میانگین چند جامعه یا چند گروه را با هم مقایسه کنیم. برای مثال فرض کنید یک شرکت با ۳ پیمان کار A, B, C قرارداد دارد و می خواهد بداند آیا میانگین یکی از مشخصات محصولات این ۳ پیمان کار با هم تفاوت دارند یا نه. برای انجام این کار از آنالیز واریانس (ANOVA) استفاده می کنیم.

الف- آنالیز واریانس یکراهه:

در این بخش می خواهیم تأثیر عامل A با a سطح را بر متغیر پاسخ y بسنجیم:

$$y_{ij} = \underbrace{\mu + \tau_i}_{\mu_i} + \varepsilon_{ij} \qquad i = 1, \dots, a \qquad j = 1, \dots, n$$

$$\begin{cases} H_0 : \mu_1 = \dots = \mu_a \\ H_1 : \text{O.W} \end{cases} \quad \text{یا} \quad \begin{cases} H_0 : \tau_1 = \dots = \tau_a \\ H_1 : \text{O.W} \end{cases}$$

مثال: فرض کنید می خواهیم تأثیر درجه حرارت محیط (در ۳ سطح) را بر سرعت واکنش (متغیر پاسخ) بسنجیم.

درجه حرارت	10	20	30
مشاهدات مربوط به پاسخ	14	18	20
	16	17	24

برای وارد کردن داده ها به طریق زیر عمل می کنیم:

```
>y = C(14, 16, 18, 17, 20, 24)
>A = C("10", "10", "20", "20", "30", "30")
>M = data.frame(x, A)
```

برای انجام آزمون $H_0: \mu_{10} = \mu_{20} = \mu_{30}$ از تابع aov( ) به صورت زیر استفاده می‌کنیم.

اسم دلخواه

```
>AN = aov ( y ~ A , data = M )
>Summary (AN)
```

منبع تغییرات	درجه آزادی	SS	MS	F-Value	P-Value	خروجی
A	a-1	SSA	MSA	MSA / MSE		
خطا	N-a	SSE	MSE			

تمرین: یک تولید کننده کاغذ می‌خواهد اثر استفاده از مواد اولیه A، B، C، D را بر روی مقاومت کششی محصولا تش بررسی نماید، بدین منظور خونه های 6 تایی را با استفاده از هریک از مواد اولیه مورد بررسی قرار می‌دهد که بعد از اندازه‌گیری نتیجه زیر حاصل می‌شود:

ماده اولیه	نتایج مشاهده‌ها					
A	14	18	19	17	16	18
B	7	8	15	11	9	10
C	19	25	22	23	18	20
D	12	17	13	18	19	15

بررسی کنید آیا مقاومت کششی کاغذها تحت تاثیر نوع مواد اولیه است یا؟

همچنین می‌توان آنالیز واریانس را با استفاده از پکیج Car و استفاده از تابع Anova( ) به صورت زیر انجام داد.

```
> S=lm ( x~A , M )
> library (Car)
> Anova(S)
```

* این دستور مدل خطی برای متغیر پاسخ (y) در برابر متغیر توضیحی (A) برازش می‌کند.

آزمون توکی (Tukey)

اگر فرض H_0 رد شود و بخواهیم بررسی کنیم کدام یک از سطوح عامل بوده که تفاوت داشتند، از آزمون توکی استفاده می‌کنیم.

برای انجام این آزمون در R، از تابع TukeyHSD( ) استفاده می‌کنیم.

```
>AN= aov ( y ~ A , data = M )
> Summary (AN)
```

→ if P_Value < α ⇒ RH_0 ⇒ $\exists i$ s.t $\tau_i \neq 0$

```
> TukeyHSD(AN)
```

بررسی فرض برابری واریانس‌ها با استفاده از آزمون بارتلت:

$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2$

```
>bartlett.test( y ~ A , data = M )
```

یکی از فرضیات برای انجام آزمون آنالیز واریانس یک طرفه، بررسی فرض برابری و همگونی واریانس‌ها در تمام سطوح است که این آزمون توسط تابع bartlett.test( ) انجام می‌گیرد.

```
>bartlett.test( y ~ A , data = M )

        Bartlett test of homogeneity of Variance

Bartlett's K-squared =    , df=    , P_Value =
```

ب- آنالیز واریانس دو راهه:

آثر دو عامل A و B بروی پاسخ تأثیرگذار باشند و این دو عامل متقاطع باشند یعنی هر سطح B در هر سطح A رخ دهد.

فرض کنید در مثال قبل عامل B معیار کارایی در دو سطح L و U باشد و داریم:

		10	20	30
L		14	18	20
		10	17	22
U		16	17	24
		18	19	27

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ij}$$

$i = 1, \dots, a$ $\quad j = 1, \dots, b$

$k = 1, \dots, n$

منبع	df	SS	MS	F	P-Value
A	$a-1$	SSA			
B	$b-1$	SSB			
A*B	$(a-1)(b-1)$	SSAB			

$H_0: \tau_1 = \dots = \tau_a = 0$

$H_0: \beta_1 = \dots = \beta_b = 0$

$H_0: \tau\beta = 0$

وارد کردن داده‌ها همانند قبل انجام می‌شود. (فرض کنید داده‌ها در یک dataframe با نام AM ذخیره شده باشند.)

```
> TA = aov(y ~ A + B + A:B, AM)
> Summary(TA)
```