Varyans Analizi

Regresyon modelinin anlamlılığını test etmek için **varyans analizi** yaklaşımı da kullanılabilir. Bu yaklaşım, yanıt değişkenindeki toplam değişkenliğin parçalanmasına dayanmaktadır.

$$y_i - \bar{y} \ = \ (\hat{y}_i - \bar{y}) \ + \ (y_i - \hat{y}_i) \qquad \textbf{(1.28)}$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i)$$

$\sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$ (özellik 1, Kesim 1.2.2) ve $\sum_{i=1}^{n}\hat{y}_i e_i = 0$ (özellik 5, Kesim 1.2.2) olmak üzere;

$$2\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = 2\sum_{i=1}^{n}\hat{y}_i(y_i - \hat{y}_i) - 2\bar{y}\sum_{i=1}^{n}(y_i - \hat{y}_i)$$

$$= 2\sum_{i=1}^{n}\hat{y}_i e_i - 2y\sum_{i=1}^{n}e_i = 0$$

olarak bulunur. Böylece aşağıdaki ifade elde edilir.

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 \ = \ \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad \textbf{(1.29)}$$

Elde edilen bu ifade, **regresyon modeli için temel varyans analizi tanımıdır.** Genelde aşağıdaki gibi yazılır:

$$SS_T \ = \ SS_R + SS_{\mathrm{Res}} \qquad \textbf{(1.30)}$$

Bu eşitlikte yer alan SS_T, gözlemlerdeki toplam değişkenliği ölçen, gözlemlerin **düzeltilmiş kareler toplamı**, SS_{R} $(SS_R = \hat{\beta}_1 S_{xy})$, **regresyon ya da model kareler toplamı** ve SS_{Res} , **artık ya da hata kareler toplamı**'dır.

- **Serbestlik derecesinin analizi :**

  1) Toplam değişim, SS_T, $df_T = n-1$ **serbestlik derecesine** sahiptir. ($y_i - \bar{y}$ sapmaları üzerindeki $\sum_{i=1}^{n}(y_i - \bar{y})$ kısıtlaması nedeniyle bir serbestlik derecesi kaybedilmiştir.)

2) Model ya da regresyon kareler toplamı, SS_R ise $df_R = 1$ **serbestlik derecesine** sahiptir. (SS_R tamamen bir parametre yani $\hat{\beta}_1$ tarafından belirlenmektedir.)

3) Artık ya da hata kareler toplamı, SS_{Res}, $df_{\mathrm{Res}} = n-2$ serbestlik derecesine sahiptir. ($\hat{\beta}_0$ ve $\hat{\beta}_1$'nın kestirilmesi sonucu $y_i - \hat{y}_i$ sapmaları üzerinde kısıtlamalar yapılmıştır.)

Serbestlik derecelerinde toplama özelliği vardır :

$$\begin{aligned} df_T &= df_R + df_{\mathrm{Res}} \\ n-1 &= 1 + (n-2) \end{aligned} \tag{1.31}$$

$H_0 : \beta_1 = 0$ hipotezini test etmek için **varyans analizi F testi** kullanılmaktadır. F istatistiği,

$$F_0 = \frac{SS_R / df_R}{SS_{\mathrm{Res}} / df_{\mathrm{Res}}} = \frac{SS_R / 1}{SS_{\mathrm{Res}} / (n-2)} = \frac{MS_R}{MS_{\mathrm{Res}}} \tag{1.32}$$

ile $F_{1,\,n-2}$ dağılımına sahiptir.

Kareler ortalamalarının beklenen değerleri,

$$E(MS_{\mathrm{Res}}) = \sigma^2 \quad , \quad E(MS_R) = \sigma^2 + \beta_1^2 S_{xx}$$

olmak üzere eğer F_0 gözlenen değeri büyükse o zaman eğimin büyük olasılıkla $\beta_1 \neq 0$ olacağını göstermektedir. Ayrıca $\beta_1 \neq 0$ olduğunda F_0, 1 ve (n-2) serbestlik dereceli aşağıdaki **merkezi olmama** parametresine sahip merkezi olmayan bir F dağılımına sahiptir.

$$\lambda = \frac{\beta_1^2 S_{xx}}{\sigma^2}$$

Bu merkezi olmama parametresi, $\beta_1 \neq 0$ iken gözlenen F_0 değerinin büyük olması gerektiğini gösterir. $F_0 > F_{\alpha,1,\,n-2}$ ise $H_0 : \beta_1 = 0$ reddedilir.

TABLO 1.4 Regresyonun Anlamlılığını Test Etmek İçin Varyans Analizi

Değişim Kaynağı	**Kareler Toplamı**	**Serbestlik Derecesi**	**Kareler Ortalaması**	F_0
Regresyon	$SS_R = \hat{\beta}_1 S_{xy}$	1	MS_R	$MS_R \,/\, MS_{\mathrm{Res}}$
Artık	$SS_{\mathrm{Res}} = SS_T - \hat{\beta}_1 S_{xy}$	n-2	MS_{Res}	
Toplam	SS_T	n-1		

Örnek 1.4 Roket Yakıtı Verileri

Roket yakıtı verileri için Örnek 1.1'de geliştirilen modelde regresyonun anlamlılığı test edilmek istensin.

$$\text{Uydurulan model : } \hat{y} = 2627.82 - 37.15x$$

$SS_T = 1,693,737.60$ ve $S_{xy} = -41112.65$ olmak üzere regresyon kareler toplamı,

$$SS_R = \hat{\beta}_1 S_{xy} = (-37.15)(-41112.65) = 1527334.95$$

$F_0 = 165.21 > F_{0.01,1,18} = 8.29$ olduğundan $H_0 : \beta_1 = 0$ hipotezi reddedilir.

- **t Testi İle İlgili Daha Fazla Bilgi**

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{\sqrt{MS_{\mathrm{Res}} / S_{xx}}} \tag{1.33}$$

olmak üzere,

$$t_0^2 = \frac{\hat{\beta}_1^2 S_{xx}}{MS_{\mathrm{Res}}} = \frac{\hat{\beta}_1 S_{xy}}{MS_{\mathrm{Res}}} = \frac{MS_R}{MS_{\mathrm{Res}}} \tag{1.34}$$

denklemi, Denklem (1.32)'deki varyans analizinin F_0 değerine eş değerdir. Roket yakıtı örneğinde, $t_0 = -12.5$ dolayısıyla $t_0^2 = (-12.5)^2 = 165.12 = F_0 = 165.21$ olarak bulunur.

***Genel olarak, f serbestlik derecesine sahip t raslantı değişkeninin karesi, payında ve paydasında sırasıyla 1 ve f sayıda serbestlik derecesine sahip bir F raslantı değişkenidir.

Regresyon bilgisayar programları hem Tablo 1.4'teki varyans analizini hem de t istatistiğini vermektedir.

TABLO 1.5 Roket Yakıtı Regresyon Modeli İçin ANOVA

Değişim Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	F_0	P-değeri
Regresyon	1527334.95	1	1527334.95	165.21	1.66×10^{-10}
Artık	166402.65	18	9244.59		
Toplam	1693737.60	19			

$\beta_1 = 0$ kararını vermek, yalnızca t ya da F istatistiğiyle desteklenen önemli sonuçlardan biridir. Eğimin istatistiksel olarak sıfırdan farklı olmaması, x ve y'nin ilişkili olmadığı anlamına gelmemektedir. Bu durum, x ve y arasındaki bu ilişkinin ölçme sürecinin değişkenliği ya da x'in değer aralığının uygun olmaması nedeniyle belirsizleştiği anlamına gelebilir.

<u>Ödev</u>

Rasgele seçilen yedi şirket için gelir ile reklam masrafı arasında doğrusal bir regresyon olduğu varsayılsın.

Reklam Masrafı (x)	1	2	4	6	10	14	20
Gelir (y)	19	32	44	40	52	53	54

a) Doğrusal regresyon modeli için EKK yardımı ile β_0, β_1 parametrelerinin tahminlerini bulunuz.

b) ANOVA tablosunu oluşturunuz ve modelin anlamlılığını test ediniz.