**9.Hafta**

## Uyumun İyiliği

Regresyon doğrusu uyumundan ne anlıyoruz? Bilindiği üzere örnek regresyon doğrusu bağımlı değişkenin tahmini değerlerinden geçer. Peki, bu tahmin edilen değerlerin gözlemlenen değerlere göre konumu nedir? Bağımlı değişkenin tahmin edilen değerlerinden geçen regresyon doğrusu bağımlı değişkenin gözlemlenen (teorik) değerler ne kadar yakınsa o derece uyumludur. Dolayısıyla, regresyon doğrusunun verilere uyumunu veren ölçülerin neler olduğu, hangisinin niçin tercih edildiği önemlidir.

### Uyumun İyiliğinin Ölçülmesi

Örneklem verileri kullanılarak tahmin edilen regresyon modelinin belirlediği regresyon doğrusu, serpilme diyagramında gözlemlenen değerlerin arasından geçmektedir. Tahmin edilen modelin başarısı açısından regresyon doğrusunun verilere ne kadar yakın olduğu önemlidir, dolayısıyla regresyon doğrusunun veriler uyumunun ölçüsüne gerek duyarız. Bunlar belirginlik katsayısı, tahminin standart hatası ve genelleştirilmiş $r^2$ dir.

Uyumun iyiliğinin araştırılmasında bilgi kriterlerinden de yararlanmak mümkündür. Bilgi kriterlerinin başlıcaları *Akaike bilgi kriteri* (AIC), *Schwarz-Bayesian bilgi kriteri* (BIC) ve *Hannan- Quinn bilgi kriteri* (HQIC)'dir.

1. **Belirginlik Katsayısı**

Bilindiği üzere regresyon modelindeki ($Y_i = \beta_0 + \beta_1 X_i + u_i$) bağımsız değişken $X_i$ , "açıklayıcı değişken" olarak da adlandırılmaktadır. Bunun nedeni ise $Y_i$'deki değişimin, $X_i$'deki değişim ile "açıklanacağı" varsayılmakta, tahmin problemine bağlı olarak bağımlı değişken $Y_i$'deki değişimin mümkün olduğunca büyük bir kısmını $X_i$ 'in açıklaması arzu edilmektedir.

$Y_i$ 'deki açıklanan değişimin bir ölçümünü geliştirmek için, $Y_i$ "açıklanabilir" ve "açıklanamayan" bileşenlerine ayrılabilir. Bağımlı değişkenin

$$Y_i = E(Y_i) + u_i$$

ile gösterimindeki birinci unsur $E(Y_i) = \beta_0 + \beta_1 X_i$ , $Y_i$'nin açıklanabilir "sistematik" bileşeni, ikinci unsur $u_i$ ise $Y_i$ 'nin açıklanamayan "rassal, sistematik olmayan" bileşenidir. Bu parçalardan $u_i$ gözlemlenemez iken, bilinmeyen parametreler $\beta_0$ ve $\beta_1$ 'yi örnek verilerinden tahmin ederek , $Y_i$ 'nin değerini benzer şekilde aşağıdaki gibi ayrıştırabiliriz.

$$Y_i = \hat{Y}_i + \hat{u}_i$$

burada $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ ve $\hat{u}_i = Y_i - \hat{Y}_i$ eşitlikleri bilinmektedir.

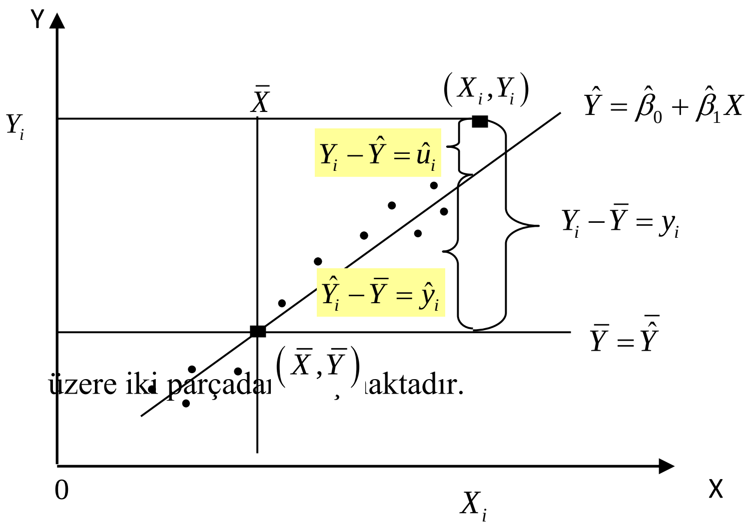
$Y_i$ 'nin yukarıdaki gibi bileşenlerine ayrılabilmesi, *EKK* özelliklerinden örnek regresyon doğrusunun "ortalamalar noktası" $(\bar{X},\bar{Y})$ noktasından geçtiği varsayımına dayanmaktadır. Buna göre $Y_i = \hat{Y}_i + \hat{u}_i$ denklemini her iki tarafından örneklem ortalaması $\bar{Y}$ 'yi çıkartarak aşağıdaki gibi yazmak mümkündür.

i. gözlem için;

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + \hat{u}_i$$

$$y_i \quad = \quad \hat{y}_i \quad + \quad \hat{u}_i$$

Aşağıdaki şekilde görüldüğü üzere, $Y_i$ ile ortalama değeri $\bar{Y}$ arasındaki fark $(Y_i - \bar{Y})$; regresyon modeli ile "açıklanan" $(\hat{Y}_i - \bar{Y})$ ve regresyonla açıklanamayan ( $Y_i - \hat{Y}_i = \hat{u}_i$ ) olmak üzere iki parçadan oluşmaktadır.




Şekil : $Y_i$'nin açıklanan ve açıklanamayan bileşenleri

Böylece *i*. gözlem için $Y_i$ 'deki örneklem değişimi

$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + \hat{u}_i$ veya $y_i = \hat{y}_i + \hat{u}_i$ ile gösterilir.

$Y_1, Y_2, \ldots Y_n$ 'e kadar bir örnekleme sahip isek, bu örneklemenin örneklem ortalaması $(\bar{Y})$ ve örneklem varyansı $\left(S_y^2 = \sum (Y_i - \bar{Y})^2 / n-1\right)$ olmak üzere iki tanımlayıcı ölçüsü olduğu istatistik derslerinden bilinmektedir. Bütün örneklem için aynı değişimin hesaplanabilmesi için $i = 1, \ldots, n$ 'e kadar örneklem değerleri $Y_i$ 'ler ile örneklem ortalaması $\bar{Y}$ arasındaki farkların kareli toplamı $\sum\left(Y_i-\bar{Y}\right)^2$ alınır ki; bu kareli toplam $\sum\left(Y_i-\bar{Y}\right)^2$, örneklem değerlerindeki *toplam değişimin* bir ölçüsüdür. ( Not: İstatistik dersinde verilen aritmetik ortalamanın özelliklerinden $\sum y_i = \sum(Y_i-\bar{Y}) = 0$ olduğu bilinmektedir. Bu nedenle $y_i$ 'lerin toplamı değil, karelerinin toplamı alınmaktadır. )

Yukarıdaki denklemin her iki tarafının karelerinin toplamı alınır

$$\sum y_i^2 = \sum(\hat{y}_i+\hat{u}_i)^2$$

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i\hat{u}_i \qquad \sum \hat{y}_i\hat{u}_i = 0 \quad \text{olduğundan}$$

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \qquad \hat{y}_i^2 = \hat{\beta}_1 x_i \text{ 'den}$$

$$\sum y_i^2 = \hat{\beta}_1^2 \sum x_i^2 + \sum \hat{u}_i^2$$

sonucuna ulaşılır. Burada;

$\sum y_i^2 = \sum(Y_i-\bar{Y})^2 =$ Toplam değişme olarak da adlandırılan Bütün Kareler Toplamı (BKT)'dır. Örneklem ortalaması etrafında $Y$'deki *toplam değişim*in bir ölçümüdür.

$\sum \hat{y}_i^2 = \sum(\hat{Y}_i-\bar{Y})^2 = \hat{\beta}_1^2\sum x_i^2 =$ Regresyona bağlı Kareler Toplamı (RKT)'dır. Aynı zamanda "regresyonla açıklanan kareler toplamı" olarak da bilinmektedir.

$\sum \hat{u}_i^2 = \sum(Y_i-\hat{Y})^2 =$ Hataya bağlı Kareler Toplamı (HKT)'dır. Aynı zamanda "açıklanamayan kareler toplamı", "kalıntı karelerinin toplamı" veya "kareli hataların toplamı" olarak bilinmektedir

Verilen kısaltmalar kullanılarak yukarıdaki eşitlik kısaca

BKT=RKT+HKT

ile de gösterilmektedir.

*Y*'deki toplam değişimin, regresyon modeli ile açıklanan ve açıklanamayan olarak iki parçaya ayrıştırılması, regresyon modeli içindeki *X* ile açıklanan değişmenin *Y*'deki değişime oranı olan, *belirginlik katsayısı* ( $r^2$ çok değişkenli regresyonda $R^2$ ) olarak bilinen bir ölçümü tanımlamamıza izin verir. Şöyle ki;

$$\sum(Y_i-\bar{Y})^2 = \sum(\hat{Y}_i-\bar{Y})^2 + \sum(Y_i-\hat{Y})^2$$

veya

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2$$

denklemlerinin her iki yanı toplam değişmeye (BKT) - $\sum y_i^2 = \sum(Y_i-\bar{Y})^2$ - bölünürse

$$1 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} + \frac{\sum \hat{u}_i^2}{\sum y_i^2}$$

sonucuna ulaşılır. Regresyon ile açıklanabilen değişmenin toplam değişmeye oranı yukarıda da belirtildiği üzere belirginlik katsayısına ( $r^2$ ) eşittir.

$$r^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\hat{\beta}_1^2 \sum x_i^2}{\sum y_i^2} = \frac{RKT}{BKT}$$

Belirginlik katsayısını Hata Kareler Toplamı (HKT) ile de göstermek mümkündür. Bunun için

$1 = r^2 + \dfrac{\sum \hat{u}_i^2}{\sum y_i^2}$ 'den $r^2$ aşağıdaki gibi yazılır.

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} = 1 - \frac{HKT}{BKT}$$

$r^2$ için yukarıda verilenlerden başka hesaplama yolları da önerilebilir. Bunun için $r^2 = \hat{\beta}_1^2 \sum x_i^2 \Big/ \sum y_i^2$ eşitliğinin pay ve paydası *n* veya *n*-1'e bölündüğünde

$$r^2 = \hat{\beta}_1^2 \frac{\sum x_i^2 / n}{\sum y_i^2 / n}$$

elde edilir. Açıktır ki; $\sum x_i^2 / n$ terimi *X* bağımsız değişkeninin koşulsuz varyansına $(S_X^2)$, $\sum y_i^2 / n$ terimi ise *Y* bağımlı değişkeninin koşulsuz varyansına $(S_Y^2)$ eşittir. Nihayet

$$r^2 = \hat{\beta}_1^2 \frac{\sigma_X^2}{\sigma_Y^2}$$

sonucuna ulaşılır.

Belirginlik katsayısı ($r^2$), örnek regresyon doğrusunun verilere uygunluğunu gösteren bir ölçüsüdür. Bu bağlamda belirginlik katsayısı, bağımlı değişkende meydana gelen değişmenin yüzde kaçının bağımsız değişken ve/veya değişkenlerdeki değişim tarafından açıklanabildiğini göstermektedir.

Belirginlik katsayısı $0 \le r^2 \le 1$ arasında yer almaktadır. $r^2$, 1'e ne kadar yakınsa örneklem değerleri ( $\hat{Y}_i$ ), $Y_i$ değerleri ile o kadar uyumlu, dolayısıyla örnek regresyon denklemi $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_i$ ana kütle regresyon denklemine ( $E(Y_i) = \beta_0 + \beta_1 X_i$ ) o nisbette yakın olacaktır.

$r^2 = 1$ ise, bağımlı değişkende meydana gelen toplam değişmenin tamamı (%100) bağımsız değişken(ler)deki değişim ile açıklanmaktadır. Söz konusu durum "*tam uyum*" olarak bilinmektedir. $r^2=0$ olması ise, eğim parametresinin $\hat{\beta}_1=0$ olması koşulana bağlıdır. Bu durumda örnek regresyon fonksiyonu $\hat{Y}_i=\hat{\beta}_0$ , regresyon doğrusu ise *X* eksenine paralel yatay bir doğru biçimindedir.

$r^2$ hangi değeri almalıdır? $r^2$ 'yi bulma ve raporlama, değişimin farklı kaynaklarının göreli büyüklükleri hakkında bilgi vermesine rağmen, belirli bir $r^2$ 'nin "yeterince büyük" olup olmadığı hakkındaki tartışmalar gereksizdir. Her ne kadar $r^2$ nin 1'e yakın olması regresyon doğrusunun verilere uygunluğu göstermekte ise de, zaman serisi verilerinin kullanıldığı modellerde trendin etkisiyle yüksek $r^2$'ye, yatay kesit verilerinin kullanıldığı modellerde ise düşük $r^2$'ye rastlamak mümkündür. Yatay-kesit verisi ile çalışılırsa $r^2$ 'nin 0.10'dan 0.40'a değerleri, çok değişkenli regresyon modelleri için bile kabul görür. Zaman boyunca çoğu kez birlikte değişme eğilimi olan zaman-serisi verisi kullanan makroekonomik analizde 0.90 ve daha yüksek bir $r^2$ değerleri elde etmek mümkündür. Dolayısıyla regresyon doğrusunun uygunluğu konusunda $r^2$ye göre yorum yaparken kullanılan veri türü dikkate alınmalıdır. Dolayısıyla $r^2$ bir istatistiktir test edilmesi gerekir. $r^2$ 'nin testi *F*-testidir. Böylece model sadece $r^2$'ye göre değil, tahminlerin işaretleri ve büyüklükleri, onların istatistiksel ve ekonomik anlamlılığı, tahminlerinin doğruluğu gibi faktörleri dikkate alarak değerlendirilmelidir.

$r^2$, *X* ve Y' nin ölçü birimine tabi değildir. Bu özelliğiyle verilerin regresyon doğrusuna uyumunu gösteren diğer ölçü birimi olan tahminin standart hatasına üstünlük sağlamaktadır.

Örnek kütle için $r^2$ ile gösterilen belirginlik katsayı, ana kütle için $\rho^2_{YX}$ ile gösterilir. Şimdi belirginlik katsayından belirsizlik katsayısına geçilecektir. Bunun için,

$$r^2=\frac{\sum \hat{y}_i^2}{\sum y_i^2} \text{'den} \qquad \sum \hat{y}_i^2 = r^2 \sum y_i^2$$

Ve yine

$$r^2=1-\frac{\sum \hat{u}_i^2}{\sum y_i^2} \text{'den} \quad \sum \hat{u}_i^2=(1-r^2)\sum y_i^2$$

eşitlikleri yazılabilir. Böylece

$$\sum y_i^2=\sum \hat{y}_i^2+\sum \hat{u}_i^2$$

eşitliği, $r^2$ ve toplam değişme açısından yeniden aşağıdaki gibi yazılır.

$$\sum y_i^2 = r^2\sum y_i^2+(1-r^2)\sum y_i^2$$

Buradaki $(1-r^2)$ belirsizlik katsayısıdır ve toplam değişmenin ne kadarının regresyonda yer almayan değişkenler tarafından açıklanabildiğini gösterir. $r^2$, 1'e yaklaştıkça, $1-r^2$ ise 0'a yaklaştıkça regresyon doğrusunun verilere uyumu artacaktır.

Basit doğrusal regresyon modelinde belirginlik katsayısı ($r^2$) ve basit korelasyon katsayısı ($r_{XY}$) arasında ilişki vardır. Örneklem veri değerleri $X_i$ ile $Y_i$ arasındaki doğrusal ilişkinin varlığını gösteren örneklem korelasyon katsayısının karesi $r_{XY}^2$, basit bir regresyon modelinde $r^2$'ye cebirsel olarak eşittir ($r_{XY}^2 = r^2$). Böylece basit regresyonda belirginlik katsayısından korelasyon katsayısına geçilebilmektedir.

**İspat**

Belirginlik katsayısı denklemi

$r^2 = \dfrac{\hat{\beta}_1^2 \sum x_i^2}{\sum y_i^2}$'da $\hat{\beta}_1$ yerine $\hat{\beta}_1 = \dfrac{\sum x_i y_i}{\sum x_i^2}$ eşitliği yazılırsa

$r^2 = \dfrac{\left(\sum x_i y_i / \sum x_i^2\right)^2 \sum x_i^2}{\sum y_i^2}$ elde edilir, sadeleştirme yapılırsa

$$r^2 = \frac{\left(\sum x_i y_i\right)^2}{\sum x_i^2 \sum y_i^2}$$

sonucuna ulaşılır ki; bu belirginlik katsayısıdır. Yukarıdaki denklemden anlaşılmaktadır ki, basit regresyon modeli için hesaplanan belirginlik katsayısı ($r^2$) korelasyon katsayısının karesine eşittir, dolayısıyla belirginlik katsayısının kare kökü basit korelasyon katsayısı verir.

$\sqrt{r^2} = r_{XY}$

$$r_{XY} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2}\sqrt{\sum (Y_i - \bar{Y})^2}} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}}$$

Burada üzerinde durulması gereken nokta belirginlik katsayısı $0 \le r^2 \le 1$ iken, korelasyon katsayısı $-1 \le r \le 1$ değerleri arasında yer alır. Belirginlik katsayısından korelasyon katsayısına geçerken değişken arasındaki ilişkinin yönünü gösteren korelasyon katsayısı işaretini, regresyon modelindeki $\hat{\beta}_1$'nin işaretinden alır.♦.

---

♦ Basit korrelasyon katsayı sıfır noktasından ve ölçekten bağımsızdır. Basit korelasyon katsayısı iki değişken arasındaki doğrusal ilişkinin yönü ve derecesimin tespiti için kullanılır. $r$=0 ise iki değişken arasında ilişki olmadığı anlamına gelmez, doğrusal ilişkinin olmadığına

**2. Tahminin Standart Hatası**

Regresyon doğrusunun verilere uygunluğunun ikinci bir ölçütü tahminin standart hatasıdır. **Örnek regresyonunun standart sapması** olarak da adlandırılan tahminin standart hatası, ana kütle hata terimi varyansı tahmininin ($\hat{\sigma}^2$) kareköküdür. Hata terimi varyansının tahmini aynı zamanda regresyondan elde edilen bilginin bir ölçüsü olduğu için karekökü olan tahminin standart hatası regresyonun verilere uyumunu gösterecektir. Tahminin standart hatası aşağıdaki gibidir.

$$\hat{\sigma} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}}$$

**Verilerin, regresyon doğrusu etrafındaki dağılmasının ölçüsü olan tahminin standart hatası**nın ($\hat{\sigma}$) büyüklüğü $Y_i$ ile $\hat{Y}_i$ arasındaki farkın büyüklüğüne bağlıdır. Bu farkın $\sum (Y_i - \hat{Y}_i)^2$ küçük çıkması, $\hat{Y}_i$' lerin verilere ($Y_i$) lere yaklaştığı, böylece regresyon doğrusunun verilere uyduğunu gösterecektir.

Ancak tahminin standart hatasının uyumun iyiliği için kullanılması, aşağıda belirtilen nedenlerden dolayı sakıncalıdır.

- Tahminin standart hatası bağımlı değişkenin ölçü birimine bağlıdır. Bağımlı değişken ton ile ölçülür ise tahminin standart hatası küçük, kg ile ölçülür ise tahminin standart hatası büyük çıkacaktır.

- Tahminin standart hatası, belirginlik katsayında ($0 \le r^2 \le 1$) olduğu gibi her durum için geçerli kesin sınırları yoktur. Regresyon doğrusunun tam uyumu durumunda, diğer bir ifade ile örnek regresyon doğrusunun $Y_i$ değerlerinden geçtiği durumda, $Y_i$ ile $\hat{Y}_i$ arasındaki fark sıfır olacaktır. Dolayısıyla tahminin standart hatası için alt sınır sıfırdır. Diğer uç bir durum ise uyumsuzluktur ancak uyumsuzlukta bir sınır getirilememektedir.

**3. Genelleştirilmiş $r^2$**

Örnek regresyon doğrusunun verilere uyumunun tespiti için bir diğer seçenek $Y_i$ 'ler ile tahmini değeri $\hat{Y}_i$ 'ler arasındaki ilişkinin saptanmasıdır. Bir çok durumda $r^2_{Y\hat{Y}}$ olarak da gösterilen $Y_i$ ve $\hat{Y}_i$ arasındaki kareli basit korelasyon, uyum iyiliği için geçerli bir ölçüdür.

---

işarettir. Örneğin $Y=X^2$ ilişkisinde $r=0$ dır. $r$'nin (-) veya (+) değer alması iki değişken arasındaki örneklem ortak varyansının işaretine bağlıdır.

Burada $\hat{Y}$ 'nin bulabilecek "en iyi" öngörü olduğu varsayılır. "en iyi" öngörücü, üzerinde düşünülen modele bağlı olarak değişebilir. Yani, genel bir uyum iyiliği ölçüsü veya genel $r^2$

$$r_G^2 = \left[cor(Y,\hat{Y})\right]^2 = r_{Y\hat{Y}}^2$$

veya daha açık bir gösterimle aşağıdaki gibidir.

$$r_{Y\hat{Y}}^2 = \left[\frac{\sum\left(Y_i-\bar{Y}\right)\left(\hat{Y}_i-\bar{\hat{Y}}\right)}{\sqrt{\sum\left(Y_i-\bar{Y}\right)^2\sum\left(\hat{Y}_i-\bar{\hat{Y}}\right)^2}}\right]^2 = \frac{\left(\sum y_i\hat{y}_i\right)^2}{\sum y_i^2\sum\hat{y}_i^2}$$

$Y_i$ ile $\hat{Y}_i$ arasındaki basit korelasyon katsayısı $-1 < r_{Y\hat{Y}} < 1$ değerleri arasında yer alacağı için, bu ifadenin karesi genelleştirilmiş $r_{Y\hat{Y}}^2$ , 0 ile 1 arasında değer alır, $0 < r_{Y\hat{Y}}^2 < 1$. Genelleştirilmiş $r_{Y\hat{Y}}^2$, uygunluk katsayısı olarak da bilinmektedir.

**Örnek: Satış Gelirleri İle Reklam Harcamaları**

Satış gelirleri (*Y*) ile reklam harcamaları(*X*) uygulamasında aşağıdaki veri ve ara sonuçlardan örnek regresyon denklemi $\hat{Y}_i = 1.0 + 1.2X_i$ olarak bulunmuştu.

| **Y** | **X** | $X^2$ | $XY$ | $\hat{Y}_i$ | $\hat{u}_i = Y_i - \hat{Y}_i$ | $\hat{u}_i^2$ |
|---|---|---|---|---|---|---|
| 3 | 1 | 1 | 3 | 2.2 | 0.8 | $(0.8)^2$ |
| 4 | 2 | 4 | 8 | 3.4 | 0.6 | $(0.6)^2$ |
| 2 | 3 | 9 | 6 | 4.6 | -2.6 | $(-2.6)^2$ |
| 6 | 4 | 16 | 24 | 5.8 | 0.2 | $(0.2)^2$ |
| 8 | 5 | 25 | 40 | 7.0 | 1 | $(1)^2$ |
| $\sum Y_i = 23$ $\bar{Y} = 4.6$ | $\sum X_i = 15$ $\bar{X} = 3$ | $\sum X_i^2 = 55$ | $\sum Y_i X_i = 81$ | $\sum \hat{Y}_i = 23$ | $\sum \hat{u}_i = 0$ | $\sum \hat{u}_i^2 = 8.8$ |

Şimdi belirginlik katsayısı, belirsizlik katsayısı, korelasyon katsayısı ve tahminin standart hatasını hesaplayalım. Bu uygulamada belirginlik katsayısı değişik yollardan hesaplanacaktır. Bunlardan ilki

$$r^2 = \frac{\sum(\hat{Y}_i-\bar{Y})^2}{\sum(Y_i-\bar{Y})^2} = \frac{\sum(\hat{Y}_i-\bar{\hat{Y}})^2}{\sum(Y_i-\bar{Y})^2} = \frac{\sum\hat{y}_i^2}{\sum y_i^2}$$

formülünün kullanılmasıdır. Buna göre formülde yer alan unsurlardan $\sum(\hat{Y}_i-\bar{Y})^2$ ve $\sum(Y_i-\bar{Y})^2$ değerlerinin hesaplanması gerekir.

| $Y_i$ | $\hat{Y}_i$ | $Y_i - \bar{Y}$ | $(Y_i - \bar{Y})^2$ | $\hat{Y}_i - \bar{Y}$ | $(\hat{Y}_i - \bar{Y})^2$ |
|---|---|---|---|---|---|
| 3 | 2.2 | -1.6 | 2.56 | -2.4 | 5.76 |
| 4 | 3.4 | -0.6 | 0.36 | -1.2 | 1.44 |
| 2 | 4.6 | -2.6 | 6.76 | 0 | 0 |
| 6 | 5.8 | 1.4 | 1.96 | 1.2 | 1.44 |
| 8 | 7.0 | 3.4 | 11.56 | 2.4 | 5.76 |
| $\sum Y_i = 23$ <br> $\bar{Y} = 4.6$ | $\sum \hat{Y}_i = 23$ | $\sum (Y_i - \bar{Y}) = 0$ | $\sum (Y_i - \bar{Y})^2 = 23.2$ | $\hat{Y}_i - \bar{Y} = 0$ | $\sum (\hat{Y}_i - \bar{Y})^2 = 14.4$ |

$$r^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = \frac{14.4}{23.2} = 0.62$$

Satışlardaki toplam değişimin % 62'si bağımsız değişken reklam harcamaları tarafından açıklanmaktadır.

Belirginlik katsayısını yukarıdaki verileri kullanarak farklı bir formül ile hesaplayalım.

$$r^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} = 1 - \frac{8.8}{23.2} = 0.62$$

Sonucun aynı olduğunu görüyoruz. Bir başka şekilde nasıl hesaplayabilirdik?

$$r^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\hat{\beta}_1^2 \sum x_i^2}{\sum y_i^2}$$

Denklemdeki toplam değişme $\sum y_i^2$, yukarıda 23.2 değerine eşit hesaplanmıştır. $\hat{\beta}_1^2$ 'yi hesaplamak gayet kolay olup $(1.2)^2$'dir. $\sum x_i^2$, $\beta_1$ tahmin edilirken 10 olarak hesaplanmıştı. Buna göre

$$r^2 = \frac{\hat{\beta}_1^2 \sum x_i^2}{\sum y_i^2} = \frac{(1.2)^2 10}{23.2} = \frac{14.4}{23.2} = 0.62$$

yine aynı sonuca ulaşılır.

Belirginlik katsayısından satışlar ile reklam harcamaları arasındaki basit korelasyon katsayısını hesaplayabiliriz.

$$r_{XY} = \sqrt{r^2} = \sqrt{0.62} = +0.79$$

Örneklem korelasyon katsayısı işaretini $\hat{\beta}_1 = 1.2$'den almıştır. $\beta_1$'in tahmini pozitif işaretli olduğu için $r_{XY}$ da pozitif işaretlilidir. Bu sonuca göre satışlar ile reklam harcamaları arasında aynı yönde (pozitif) güçlü (0.79) bir doğrusal ilişki vardır. Basit korrelasyon katsayısını $r = \sum x_i y_i \Big/ \sqrt{\sum x_i^2 \sum y_i^2}$ formülünden hesaplanarak da aynı sonuca ulaşıldığı görülebilir.

Belirginlik katsayısından, belirsizlik katsayısı $1 - r^2$'yi de hesaplamak mümkündür. Belirsizlik katsayısı $1 - r^2 = 1 - 0.62 = 0.38$'dir. Bu sonuca göre satışlardaki toplam değişmenin %38'i reklam harcamaları dışındaki değişkenler tarafından açıklanmaktadır.

Regresyon denkleminin uyumunu gösteren diğer bir ölçü tahminin standart hatasıdır. Tahminin standart hatası: $\hat{\beta}_1^2$ değerine eşittir.

Bu örnek ile ilgili sonuçlar aşağıdaki gibi raporlanır.

$\hat{Y}_i = 1.0 + 1.2X_i \quad r^2 = 0.62 \quad \hat{\sigma} = 1.71 \quad n = 5$

$Se(\beta_i)(1.795)(0.541)$

**Örnek: Tüketim Harcamaları İle Gelir**

**Örneklem 1**

Örneklem 1 için aşağıdaki ara sonuçlardan

n=10 $\sum Y = 1211$ $\sum X_i = 1700$ $\sum X_i^2 = 322000$ $\sum Y^2 = 159579$ $\sum Y_i X_i = 226020$ $\bar{X} = 170$ $\bar{Y} = 121.1$

$\hat{Y}_i = 17.29 + 0.61X_i$ regresyon denklemini tahmin etmiştik. Şimdi modelin belirginlik katsayısı, belirsizlik katsayısı, korelasyon katsayısı ve tahminin standart hatasını hesaplayacağız. Belirginlik katsayısını farklı yollarla hesaplayabileceğimizi biliyoruz. Bunlardan birini Örneklem 1 için, bir başkasını Örneklem 2 için kullanacağız. Örneklem 1 için aşağıdaki formülü kullanıyoruz.

$$r^2 = \frac{\sum \hat{y}^2}{\sum y^2} = \frac{\hat{\beta}_1^2 \sum x_i^2}{\sum y^2}$$

Öncelikle regresyon ile açıklanan değişme (RKT ) $\sum \hat{y}^2$ aşağıdaki eşitlikten hesaplamamız gerekir.

$$\sum \hat{y}^2 = \hat{\beta}_1^2 \sum x_i^2$$

Yukarıdaki eşitlikte yer alan $\hat{\beta}_1$ bilinmekte ancak $X$'deki değişme $\sum x_i^2$ bilinmemektedir. Öncelikle ara sonuçlardan $\sum x_i^2$ 'nin hesaplanması gerekir.

$$\sum x_i^2 = \sum X_i^2 - n\bar{X}^2 = 322000 - \left(10 \times 170^2\right) = 33000$$

Buradan regresyon ile açıklanan değişme:

$\sum \hat{y}^2 = \hat{\beta}_1^2 \sum x_i^2 = (0.61)^2 \times 33000 = 12279.3$ olarak hesaplanır.

Şimdi belirginlik katsayının ikinci bileşeni toplam değişmeyi (BKT) $\sum y^2$ hesaplayacağız.

Toplam değişme

$\sum y^2 = \sum Y^2 - n\bar{Y}^2$

eşitliğinden hesaplanır. Ancak yukarıdaki ara sonuçlarda gözlemlenen *Y*'lerin karelerinin toplamı ( $\sum Y_i^2$ ) verilmediği için öncelikle $\sum Y_i^2$ 'nın hesaplanması gerekir.

| **ÖRNEKLEM 1** | |
|---|---|
| $Y_i$ | $Y_i^2$ |
| 65 | 4225 |
| 80 | 6400 |
| 79 | 6241 |
| 113 | 12769 |
| 125 | 15625 |
| 115 | 13225 |
| 144 | 20736 |
| 157 | 24649 |
| 155 | 24025 |
| 178 | 31684 |
| | $\sum Y_i^2 = 159579$ |

$\sum Y_i^2 = 159579$ hesaplandığına göre

$$\sum y^2 = \sum Y^2 - n\bar{Y}^2 = 159579 - 10(121.1)^2 = 12926.9$$

Buradan belirginlik katsayısı

$$r^2 = \frac{\hat{\beta}_1^2 \sum x_i^2}{\sum y^2} = \frac{12279.3}{12926.9} = 0.95$$

Bu sonuca göre tüketimdeki toplam değişmenin %95' i gelir ile açıklanmaktadır. Belirsizlik katsayısı;

$$1 - r^2 = 1 - 0.95 = 0.05$$

Tüketimdeki toplam değişmenin %5'i gelir dışındaki başka değişkenler ile açıklanmaktadır. Gelir ile tüketim arasındaki basit korelasyon katsayısı:

$$r = \sqrt{r^2} = \sqrt{0.95} = +0.97$$

Korelasyon katsayısı işaretini $\hat{\beta}_1$ 'in işaretinden almıştır. Gelir ile tüketim arasında aynı yönde güçlü doğrusal bir ilişki vardır.

Regresyon doğrusunun verilere uyumunu gösteren diğer bir ölçü tahminin standart hatasını hesaplayalım.

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}} =$$

Burada kalıntı kareler toplamını bilmiyoruz. Ancak

$$\sum y^2 = \sum \hat{y}^2 + \sum \hat{u}_i^2$$

eşitliğinden hesaplayabiliriz. Buna göre yukarıda hesapladığımız değerleri yerine koyarsak;

$$12926.9 = 12279.3 + \sum \hat{u}_i^2$$

Buradan

$$\sum \hat{u}_i^2 = 12926.9 - 12279.3 = 647.6$$

olarak hesaplanır ve tahminin standart hatası aşağıdaki gibidir.

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}} = \sqrt{\frac{647.6}{10-2}} = \sqrt{80.95} = 8.997$$

Örneklem 1 için elde edilen sonuçlar aşağıdaki gibi raporlanır.

$$\hat{Y}_i = 17.29 + 0.61X_i \qquad r^2 = 0.95 \qquad \hat{\sigma} = 8.997 \qquad n = 10$$
$$Se(\beta_i)\,(8.72)(0.049)$$

*Not:*

$r^2 = \frac{\sum \hat{y}^2}{\sum y^2} = \frac{\hat{\beta}_1^2 \sum x_i^2}{\sum y^2}$ eşitliği aynı zamanda gözlemlenen değerler ile $r^2 = \frac{\sum \hat{y}^2}{\sum y^2} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$ ile gösterilmektedir. Satış örneğinde olduğu gibi $\frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$ eşitliğini kullanarak belirginlik katsayısını hesaplayabilirsiniz.

**Örneklem 2.**

Örneklem 2 için aşağıdaki ara sonuçlardan

n=10 $\sum Y = 1207$ $\sum X_i = 1700$ $\sum X_i^2 = 322000$ $\sum Y^2 = 161441$ $\sum Y_i X_i = 226800$ $\bar{X} = 170$

$\bar{Y} = 120.7$

$\hat{Y}_i = 9.37 + 0.65X_i$ regresyon denklemi tahmin edilmişti. Örneklem 2 için tahmin edilen modelin belirginlik katsayısı, belirsizlik katsayısı, korelasyon katsayısı ve tahminin standart hatasını hesaplayacağız. Örneklem 2 için aşağıdaki formülü kullanıyoruz.

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$$

Yine burada toplam değişme $\sum y_i^2$ ve regresyonla açıklanamayan değişme hata kareleri toplamı ($\sum \hat{u}_i^2$) bilinmemektedir. Öncelikle toplam değişmeyi ($\sum y_i^2$) hesaplayalım.

$$\sum y^2 = \sum Y^2 - n\bar{Y}^2 =$$

| **ÖRNEKLEM 2** | |
|---|---|
| $Y_i$ | $Y_i^2$ |
| 55 | 3025 |
| 74 | 5476 |
| 90 | 8100 |
| 103 | 10609 |
| 107 | 11449 |
| 135 | 18225 |
| 144 | 20736 |
| 160 | 25600 |
| 189 | 35721 |
| 150 | 22500 |
| | $\sum Y_i^2 = 161441$ |

$$\sum y^2 = \sum Y^2 - n\bar{Y}^2 = 161441 - 10(120.7)^2 = 15756.1$$

Formülde diğer bilinmeyen kalıntı kareler toplamı $\sum \hat{u}_i^2$ 'dır. Ancak $\sum \hat{u}_i^2$ aşağıdaki eşitlikten hesaplanabilmesi için regresyonla açıklanan değişmenin $\sum \hat{y}^2$ hesaplanması gerekir.

$$\sum y^2 = \sum \hat{y}^2 + \sum \hat{u}_i^2 =$$

Regresyonla açıklanan değişme ise aşağıdaki gibidir.

$$\sum \hat{y}^2 = \hat{\beta}_1^2 \sum x_i^2 =$$

Ancak burada da X deki değişmenin yukarıda verilen verilerden hesaplanması gerekecektir.

$$\sum x_i^2 = \sum X_i^2 - n\bar{X}^2 = 322000 - \left(10 \times 170^2\right) = 33000$$

*Önemli not*: Her iki örneklemde de X'ler sabit olduğu için X'deki değişmenin aynı olduğuna dikkat edin.

Buradan hesaplamaları yapabiliriz. Regresyonla açıklanabilen değişme aşağıda hesaplanmıştır.

$$\sum \hat{y}^2 = \hat{\beta}_1^2 \sum x_i^2 = (0.65)^2 \times 33000 = 13942.5$$

Toplam değişmenin bileşenlerinden regresyon ile açıklanamayan değişmeyi bulabiliriz.

$$\sum y^2 = \sum \hat{y}^2 + \sum \hat{u}_i^2$$

$$15756.1 = 13942.5 + \sum \hat{u}_i^2$$

$$\sum \hat{u}_i^2 = 15756.1 - 13942.5 = 1813.6$$

Şimdi belirginlik katsayısını hesaplayabiliriz.

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} = 1 - \frac{1813.6}{15756.1} = 0.88$$

Böylece tüketimdeki toplam değişmenin %88' gelir değişkeni ile açıklanmaktadır. Belirsizlik katsayısı;

$$1 - r^2 = 1 - 0.88 = 0.12$$

Tüketimdeki toplam değişmenin %12'si gelir dışındaki modele girmeyen değişkenler ile açıklanmaktadır. Gelir ile tüketim arasındaki basit korelasyon katsayısı

$$r = \sqrt{r^2} = \sqrt{0.88} = +0.94$$

Gelir ile tüketim arasında aynı yönlü güçlü doğrusal bir ilişki vardır. Tahminin stadart hatası:

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}} = \sqrt{\frac{1813.6}{10-2}} = \sqrt{226.7} = 15.057$$

olarak hesaplanır.

*Not:* $r^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$ eşitliği bilindiğine göre, gözlemlenen verilerden elde ettiğiniz sonuçları $r^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2}$ denkleminde yerine koyar ve hesaplamaları yaparsanız aynı sonuca ulaşırsınız.

Örneklem 2 için elde edilen sonuçlar aşağıdaki gibi raporlanır.

$$\hat{Y}_i = 9.37 + 0.65X_i \qquad r^2 = 0.88 \qquad \hat{\sigma} = 15.057 \qquad n = 10$$
$$Se(\beta_i)\,(14.02)(0.08)$$

Yukarıda aynı ana kütleden çekilen iki örneklemden siz olsanız hangisi tercih edersiniz? Hiç kuşkusuz cevabınız 1. Örneklem olmalı. Çünkü belirginlik katsayısı 2. Örnekleme göre daha büyük, belirsizlik katsayısı ve tahminin standart hatası ise daha küçüktür.

**Örnek: Gıda harcaması- Gelir**

$\hat{Y}_i = 83,4160 + 10,21X_i \qquad \hat{\sigma}^2 = 8013,29$

$Se(\hat{\beta}_i)\ (43,410) \quad (2,093)$

$$\sum (Y_i - \bar{Y})^2 = \sum y_i^2 = 495132,160$$

$$\sum (Y_i - \hat{Y}_i)^2 = \sum \hat{u}_i^2 = 304505,176$$

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} = 1 - \frac{304505,176}{495132,160} = 0,385$$

$$r^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} =$$

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2$$

$$\sum \hat{y}_i^2 = \sum y_i^2 - \sum \hat{u}_i^2$$

$$\sum \hat{y}_i^2 = 495132,160 - 304505,176$$

$$\sum \hat{y}_i^2 = 190626,984$$

$$r^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{190626,984}{495132,160} = 0,385$$

Gıda harcamasındaki toplam değişimenin %38.5'i gelir ile açıklanmaktadır. Belirsizlik katsayısı $1 - r^2 = 1 - 0.385 = 0,615$ 'e eşittir. Dolayısıyla gelir harcamalarındaki toplam değişimenin yaklaşık % 62'si gelir değişkeni dışındaki değişkenler tarafından açıklanmaktadır. Tahminin standart hatası $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{8013,29} = 89,52$ olarak hesaplanır. Gıda harcamaları ile gelir arasındaki basit korelasyon katsayısı ise $r = \sqrt{r^2} = \sqrt{0,385} = +0,62$ 'e eşittir.