YYÜZF DERGİSİ 1991 1/2 (155-162)

$\bar{x}$'in μ ve $\hat{\beta}_1$'in β_1 İÇİN EN İYİ MİNİMUM VARYANSLI SAPMASIZ TAHMİNLEYİCİ (EMVST) OLMA DURUMU ÜZERİNE BİR TARTIŞMA

H.OKUT[1] O.KARACA[2]

(DERLEME)

ÖZET

İki veya daha fazla değişken arasındaki matematiksel ifadeye regresyon denklemi denir. $\hat{Y}_i=\hat{\beta}_o+\hat{\beta}_1X_1$ regresyon denkleminde, $\hat{Y}_i$, $\hat{\beta}_o$ ve $\hat{\beta}_1$ sırasıyla Y_i, β_o ve β_1'in tahminleyicileridir. Bu nedenle $\hat{Y}_i=\hat{\beta}_o+\hat{\beta}_1x_1$ hatasızdır. Eğer parametre vektörüne $\theta(\beta_1,\beta_2....\beta_n)$ dersek, bunun tahminleyiciside $\hat{\theta}(\hat{\beta}_1,\hat{\beta}_2...\hat{\beta}_n)$'dir. Aynı şekilde örnek ortalaması $\bar{x}$ olursa, temsil edeceği populasyonun ortalaması μ olur. İyi bir tahminleyici sapmasız, kararlı, yeterli ve etkin olmalıdır. En iyi tahminleyici belirlemede bir çok metod geliştirilmiştir. Bunların içinde yaygın olarak en küçük kareler metodu ile en yüksek olasılık metodları kullanılmaktadır. Bu makalede $\bar{x}$'in, μ için ve $\hat{\beta}_1$-'in β_1 için en iyi minimum varyanslı sapmasız tahminleyicisi (EMVST) olduklarına dair metod ve teoremlerden bazıları tanıtılmıştır.

A DISCUSSION ON THE UNBIASED ESTIMATOR WITH THE BEST MINIMUM VARIANCE OF $\bar{X}$ FOR μ AND $\hat{\beta}_1$ FOR β_1

SUMMARY

1-Yüz.Yıl Üniv. Ziraat Fak. Zootekni Bölümü, Araş.Gör.

2-Yüz.Yıl Üniv. Ziraat Fak. Zootekni Bölümü, Yard.Doç.Dr.

A mathematical statment which determins the relation among two or more than two variables is called a regression equation. In the regression equation $\hat{y}_i=\hat{\beta}_o+\hat{\beta}_1x_1$, $\hat{y}_i$, $\hat{\beta}_o$ and β_1 denote the estimotors of y_i, B_o and B_1 respectively. Therefore the equation $\hat{y}_i=\hat{\beta}_o+\hat{\beta}_1x_1$ is a statistic and it has no error. If we denote the parameter vector by $\theta(\beta_1, \beta_2, \dots \beta_n)$ than its estimator is given as $\hat{\theta}(\hat{\beta}_1, \hat{\beta}_2, \dots \beta_n)$. Similarly, if the sample means is $\bar{x}$, the means of population represented by them would be μ. A good estimator must be unbiased, consistent, sufficient and efficient. The best estimator has developed different kinds of methods. Among them, least square and maximum likelihood methods are used commenly. In this paper, some methods and theorems that proves the best minimum variance unbiased estimator of $\bar{x}$ for μ and $\hat{\beta}_1$ for β_1 are represented.

GİRİŞ

Basit olarak iki veya daha fazla değişken arasındaki ilişkinin matematik kurallara göre açıklanmasına **"Regresyon Denklemi"** denir (1). Bu ilişkiler doğrusal (Linear) ve doğrusal olmayan (Non-Linear) karaktere sahip olabilirler. Değişkenlerden bir tanesi bir veya daha fazla değişkenin fonksiyonu şeklinde olur. Bu fonksiyonel ilişkinin modeli parametrik katsayılarla ifade edilmiş ise bu katsayıların hatasız olduğunu kabul ediyoruz (1). Yani $\hat{y}_i=\hat{\beta}_o+\hat{\beta}_1x_i$ şeklinde yazılan bir modelde değişkenler arasında linear ve hatasız bir ilişki vardır ve x ile y değerleri (noktaları) aynı doğru üzerinde dizilirler. Ancak bu yazılan model matematiksel bir denklemdir. İstatistiki bir denklem yazılınca değişkenler arasındaki stokastik ilişki dikkate alındığı için modele hata teriminde ilave edilmesi gereklidir.

$y_i=\beta_o+\beta_1x_i+E_i$, $y_i=\beta_o+\beta_1x_1+\dots+\beta_nx_n+E_i$ veya $y_i=\beta_o+\beta_1x_1+\beta_2x_2+\dots+$

$\beta_n x_n + \beta_{11} x_1^2 + \ldots + \beta_{nn} x_n^2 + \beta_{12} x_1 x_2 + \ldots + E_i$ şeklinde yazılabilen regresyon modellerini hata (E_i) terimi ilave edildiği taktirde, populasyonu temsil eden bu denklemlerin oluşmasında, değişkenler arasında stokastik ilişkilerde dikkate alındığını ifade eder.

Eğer modelde yer alan parametreleri populasyon düzeyinde belirleme imkanına sahip değilsek (hemen hemen bu durum olanaksız) bunların değerini istatistiki olarak tahmin ederiz. Eğer parametre vektörüne $\theta(\beta_1, \beta_2, \ldots \beta_n)$ dersek, bunun tahminleyicisi de $\hat{\theta}(\hat{\beta}_1, \hat{\beta}_2, \ldots \hat{\beta}_n)$ olur. θ'nın $\hat{\theta}$ için iyi bir tahminleyici olabilmesi için (2, 3, 4, 5, 6, 7, 8):

a- Sapmasız (Unbiased) olmalı. Yani: θ'nın tahminleyicisi olan $\hat{\theta}$'nın beklenen değeri sıfır olmalı $E(\theta)=0$

b- Kararlı (Consistent) olmalı. Her bir $E_i > 0$ durumunda $\hat{\theta}$'nın θ için kararlı bir tahminleyici olabilmesi için,

$\lim_{n \to \infty} P(|\hat{\theta}-\theta| \geq E)=0$ olması gerekir.

c- Yeterlilik (Sufficiency) söz konusu olmalı. Eğer $\hat{\theta}$ şeklinde tanımlayacağımız gözlemlere ait bir fonksiyon bulunuyorsa ve $\hat{\theta}$ ifadesi örnekte, θ parametresine ait bütün bilgileri içeriyorsa bu durumda $\hat{\theta}$, θ için yeterlilik durumunu sağlar.

$$\frac{\partial L_n L}{\partial \theta} = \text{yalnız } \theta \text{ için } \hat{\theta}\text{'nın bir fonksiyonudur.}$$

d- Etkinlik (Efficiency). Tahminlemenin asıl amacı yüksek bir ihtimal ile θ tahminlemek olduğuna göre sapmasız minimum varyanslı tahminleri seçmek gerekir. Eğer bağımsız değişkenlerin birlikte yoğunluk fonksiyonu $f(x_1)f(x_2)\ldots f(x_n)$ olursa

ve $$\frac{\partial \operatorname{Ln} C}{\partial \theta} = \lambda(\theta)(\hat{\theta}-\theta)$$ ise etkin bir tahminleyici vardır. Burada

C birleşik yoğunluk, $\lambda(\theta)$, θ'nin yaklaşık fonksiyonu ve $\hat{\theta}$ x'in fonksiyonudur. Bu durumda minimum varyanslı sapmasız tahminleyici varyansı $1/\lambda(\theta)$ olur (3, 9, 10, 11).

Yukarıda ifade edilen en gerçekçi tahminleyici bulmada, hali hazırda birkaç metod kullanılmaktadır. Bunlardan en yaygın olanları:

1- Maksimum olasılık metadu,

2- En küçük kareler metodu,

3- Momentler metodu,

4- Bayesian metodudur (3, 6, 9).

Bu makalede $\bar{X}$'in μ ve $\hat{\beta}_1$'in β_1 için en iyi doğrusal (linear) sapmasız tahminleyici olma durumu esas alınarak, bu iki tahminleyicinin, parametreleri için en iyi minimum varyanslı sapmasız tahminleyici (EMVST) olma durumu tartışılacaktır.

3- $\bar{X}$'in μ İçin En İyi Minimum Varyanslı Sapmasız Tahminleyici (EMVST) Olması

θ'nın populasyon parametreler vektörü $(\theta_1 \ldots \theta_n)$ ve $\hat{\theta}$'nında bu parametrelerin en iyi (Best) sapmasız min.-varyanslı tahminleyicileri $(\hat{\theta}_1 \ldots \hat{\theta}_n)$ olduğunu düşünürsek, bunun gerçekleşmesi için;

$$v(\hat{\theta}) \geqslant \frac{(\theta)^2}{nE\left[\frac{\partial}{\partial \theta} Lnf(x_i, \theta)\right]^2} \Longrightarrow v(\theta) \geqslant \frac{1}{nE\left[\frac{a}{a\theta} Lnf(x_i)\right]^2}$$

eşitsizliğini kullanmak gerekir (11). Cramer-Rao eşitsizliği denilen bu formül normal populasyondan çekilmiş n birey içeren bir örneğe ait dağılışın fonksiyonunu tanımlamaktadır. Ancak bu eşitsizliğin geçerli olabilmesi için şu varsayımlar yapılmaktadır (9).

$X_1 \ldots X_n$ şans örneği, $f(X_i, \theta)$ dağılışından çekilmiş olsun. $\hat{\theta}_1$, $\hat{\theta}_2 \ldots \hat{\theta}_n$ θ_i'nin tahminleyicileri olursa;

i- $\frac{\partial}{\partial \theta} \int \ldots \int \prod_{i=1}^{n} f(x_i, \theta) dx_1 \ldots dx_n$ olur.

ii- $\frac{\partial}{\partial n \theta} Lnf(x_i, \theta)$ türevi mevcuttur.

iii- $\underbrace{\int \ldots \int \hat{\theta}_1 \hat{\theta}_2 \ldots \hat{\theta}_n \frac{\partial}{\partial \theta} \prod f(x_i, \theta) dx_1 \ldots dx_n}_{E(\theta)}$

iv- θ $E(\frac{\partial}{\partial \theta} Lnf(x_i, \theta)^2$

gibi varsayımlar geçerli olduğu taktirde Cramer-Rao eşitsizliği doğrudur (9, 11). Bu durumda X $N(\mu, \sigma^2)$ karakterindeki bir populasyondan çekilen n bireyli örneğin ortalaması $\bar{x}$, bunun temsil ettiği populasyonun ortalaması μ ise, örnek ortalaması ($\bar{x}$), populasyon ortalamasının en iyi (best) sapmasız min.- varyansı tahminleyici olur.

$N(\mu, \sigma^2)$ gibi normal dağılıştan çekilmiş $x_1 \ldots x_n$ şans değişkeni bulunsun. Bu durumda $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ olan bir normal dağılış gösterir. Eğer $\bar{x}$, μ'nin ESMV tahminleyicisi ise; $V(\bar{x}) = \frac{\sigma^2}{n}$ olması gerekir.

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{(X-\mu)^2}{2\sigma^2}}$$

Normal dağılışın yoğunluk fonksiyonunda π ve e değerleri matematiksel olarak sabit σ^2 ve N dağılışın parametreleri, x değeri ise $-\infty$ ile $+\infty$ arasında değerlere sahiptir. x'in herhangi bir k değerinde küçük olma alasılığı;

$\int_{-\infty}^{k} f(x)dx = \int_{-\infty}^{k} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{X-\mu}{\sigma})^2} dx$ olup $-\infty$ ile $+\infty$ arasındaki değerlerin oranı 1 olur (9, 12, 13).

$$Ln(fx) = -\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} Ln\sigma - Ln\sigma\sqrt{2\pi}$$

$Ln\sigma = 1$ olduğundan;

$$Lnf(x) = -2n6 \sqrt{2\pi} - \frac{1}{2} \left(\frac{x-\mu}{6} \right)^2$$

$\frac{\partial Lnf(x)}{\partial \mu} = \frac{x-\mu}{\sigma^2}$ olur.

$$E\left[\left(\frac{\partial Ln\ f(x_i)}{\partial \mu}\right)\right] = \frac{1}{\sigma^2}\ E\ \frac{(x-\mu)^2}{\sigma^2} \Longrightarrow \frac{1}{\sigma^2}$$ olur (9).

Cramer-Rao eşitsizliğinde,

$$V(\hat{\theta}) = \frac{(\theta)^2}{nE\left|\frac{\partial\ Ln\ f(xi;\theta)^2}{\partial\theta}\right]} = \frac{1}{\sigma^2} \Longrightarrow \frac{1}{\frac{n}{\sigma^2}} \Longrightarrow \frac{\sigma^2}{n}$$ olur.

O halde $\bar{x}$ μ'nin ESMV tahminleyicisidir (9, 11).

$Y=\beta_o-\beta_1x_1+E_i$ Doğrusal Modelde $\hat{\beta}_i$'nin β_1 İçin En İyi Minimum Varyanslı Sapmasız Tahminleyici (EMVST) Olma Durumu

$E(E_i)=0$

$V(E)=\sigma^2$ varsayımı geçerli ve $E\sim\mu(0,\sigma^2)$ dır (11). Aynı şekilde

$V(Y)=\sigma^2$ $\mu=\beta_1x+\beta_o$ olduğundan,

$$\hat{\beta}_1 = \frac{(\Sigma Wi)\ (\Sigma WiXiYi)-(\Sigma WiXi)\ (\Sigma WiYi)}{(\Sigma Wi)\ (\Sigma Wi\ Xi^2)-(\Sigma WiXi)^2}$$

$$\hat{\beta}_1 = \frac{\Sigma WiYi(Xi-\bar{X})}{\Sigma Wi(Xi-\bar{x})^2} = \frac{\Sigma WiXiYi-(\Sigma WiXi)\ (\Sigma WiXi/\Sigma Wi)}{\Sigma WiXi^2-(\Sigma WiXi)^2/\Sigma Wi}$$

Burada; $\bar{x}=\Sigma WiXi/\Sigma Wi$

$Y=\Sigma WiYi/\Sigma Wi$

Bütün i'ler için $V(Y_i)=\sigma^2$'dır. Basit olarak bu formül;

$$\hat{\beta}_1 = \frac{n\Sigma XiYi-(\Sigma Xi)(\Sigma Yi)}{n\Sigma Xi^2-(\Sigma Xi)^2} = \frac{\Sigma Yi(Xi-\bar{x})}{\Sigma(Xi-\bar{x})^2}$$ olur (1,2,5,6,8,10,14).

$(Xi-\bar{x})$ = Mi dersek

$\hat{\beta}_1 = \frac{\Sigma MiYi}{\Sigma Mi^2}$ olur ve $\frac{Mi}{\Sigma Mi^2} = k_i$ olursa $\hat{\beta}_1=\Sigma k_iYi$ olur. $\hat{\beta}_i$'nin varyansını bulacak olursak;

$V(\hat{\beta}_1)=V(k_iY_i)=\Sigma k_i^2 V(Yi)$

Hatırlanacağı gibi $V(Yi)=\sigma^2$ ve $Yi \sim (\beta_o+\beta_1 x_1;\sigma^2)$ olduğundan

$V(\hat{\beta}_1)=\sigma^2\Sigma k_i^2 = \sigma^2\Sigma(\frac{Mi}{Mi^2})^2$ olur (9, 11). $\hat{\beta}_1$'in β_1 için ESMVT ola-

bilmesi için $k_i^2=(\frac{Mi}{Mi^2})^2$'nin minimum olması gerekir. Bunu ise Gauss -Markov teoremi ile ispatlayabiliriz (11).

Ayrıca;

$$\beta_1=\frac{\Sigma(X-\bar{x})Yi}{\Sigma(X-\bar{x})^2}=LiYi \Longrightarrow Li-\frac{X-\bar{x}}{E(X-\bar{x})^2} \quad \text{olsun}$$

Bu durumda;

$\hat{\beta}_1 = MiYi$

$B_1 = LiYi$ olur ve $E(\beta_1)=\beta_1$ (sapmasız olduğundan)

$d=\hat{\beta}_1-\beta_1$ diyelim. Sapmasızlık varsayımından $E(d)=0$ olur.

$d=\Sigma MiYi-\Sigma LiYi \Longrightarrow \Sigma(Mi-Li)Yi$

$E(d)=\Sigma(Mi-Li)E(Yi) \Longrightarrow E(Yi)=\beta_o+\beta_1X_1$ olduğunda

$E(d)=E(Mi-Li)(\beta_o+\beta_1X_1)$

$E(d)=0$ olması için

$\Sigma(Mi-Li)=0$ olmalı.

$$Cov(\beta_1,d)=Cov[\Sigma LiYi,\ \Sigma(Mi-Li)Yi]$$
$$=\sigma^2[\Sigma Li(Mi-Li)]=\sigma^2\frac{1}{\Sigma(X-\bar{x})^2}[(Xi-\bar{x})(Mi-Li)] \quad 0$$

$Cov(B,\ d)=0$ olur.

$\hat{\beta}_1=\beta_1+(\hat{\beta}_1-\beta_1) \Longrightarrow (\hat{\beta}_1-\beta)=d$

$V(\hat{\beta}_1)=V(B_1)+V(d)+2Cov(B_1,d)$ 0

$V(\hat{\beta}_1)=V(\beta_1)+V(d)\geq(\beta_1)$ olduğundan ve Gauss-Markov teoremi gereği, $\hat{\beta}_1$, β_1'in ESMV tahminleyicisi olur (9, 10, 11).

Sonuç olarak ortalaması sıfır, varyansı σ^2 olan μ'nın $\mu \sim N(0, \sigma^2)$ EMVST $\bar{x}$ olup bunun ortalaması μ ve variyansı σ^2/n dir. Keza $V(\hat{\beta}_1)=V(\beta_1)+V(d)$ olduğunda ve $E(d)=0$ olduğunda $\hat{\beta}_1$, β_1'in EMVST olmaktadır.

YARARLANILAN KAYNAKLAR

1. DÜZGÜNEŞ, O., KESİCİ, T., KAVUNCU, O., GÜRBÜZ, F., Araştırma ve deneme Metodları. A.Ü.Zir.Fak.Yay.No:1021, Ders Kitabı, 295, 1987, Ankara.
2. MOSTELLER, F., TUKEY, W.J., Data Analysis And Regression. Addison-Wasley Publishing Company. 1977. California.
3. GREEN, J.R. and MARGERISON, D., Statistical Treatment Of Experimental Data. Elsevier Scientific Publishing Company. 1978, New York.
4. SOKAL, R.R. and ROHLF, J.F., Biometry. New York, 1981.
5. STEEL, G.R., TORRIE, H.J., Principles And Procedures Of Statistics. McGrow-Hill Book Company. New York, 1980.
6. METER, J. and WASSERMAN, W., Applied Linear Statistical Models Richard D. Irwin, Inc. 1974. Illinois 60430.
7. WINER, J.R., Statistical Principles In Experiments Chapman And Hill, Limited 1952, London.
9. ÖZTÜRK, A., Matematiksel İstatistik Ders Notları. 1985, İzmir.
10. DRAPPER, N.R. and SMITH, H., Applied Regression Analysis. Wiley. 1982, New York.
11. ERGEN, Ö., Uygulamalı Regresyon Ders Notları. 1989, İzmir.
12. SEZGİN, F., İstatistik Ders Notları. 1980, Erzurum.
13. DÜZGÜNEŞ, O., KESİCİ, T., GÜRBÜZ, F., İstatistik Metodları I. A.Ü.Zir.Fak.Yay:861, Ders Kitabı:229, 1983, Ankara.
14. COCHRON, G.W., COX, M.G., Experimental Design Jhon Wiley And Sons 1976, New York.