

## Yığın Hacminin Tahmini İçin Bulanık Doğrusal Regresyon Modelinde Ters Tahmin Metodu

Mustafa SEMİZ<sup>1</sup>, Aşır GENÇ<sup>2</sup>

**Özet:** Bu çalışmada yığın hacminin tahmini için farklı bir yaklaşım sunulmaktadır. Yığın hacminin tahmini için yakalama-tekrar yakalama yöntemiyle elde edilen bağımlı ve bulanık gözlemlere dayanan ters tahmin yönteminde bulanık doğrusal regresyon modelleri önerilmektedir. Bu çalışmada, iki farklı bulanık model bir örnek üzerinde tartışılmış ve bu modeller için varsayımlar belirlenmiştir.

**Anahtar kelimeler:** Bulanık veri; Doğrusal regresyon modeli; Enküçük kareler yöntemi; yakalama-tekrar yakalama yöntemi; ters tahmin yöntemi; yığın hacmi tahmini.

## Inverse Estimation Method In Fuzzy Linear Regression Model For Estimate Of Population Volume

**Abstract :** In this study, a different approach is presented for estimating population size. For estimate of population size, fuzzy linear regression models in inverse prediction methods are suggested with the data, dependent and fuzzy observations, obtained from capture-recapture sequential sampling process. In this study, two different fuzzy models are investigated on an example and then the required assumptions of these models are determined.

**Key words:** Fuzzy data; Linear regression model; Least squares method; capture-recapture; inverse prediction method; population estimate.

### Giriş

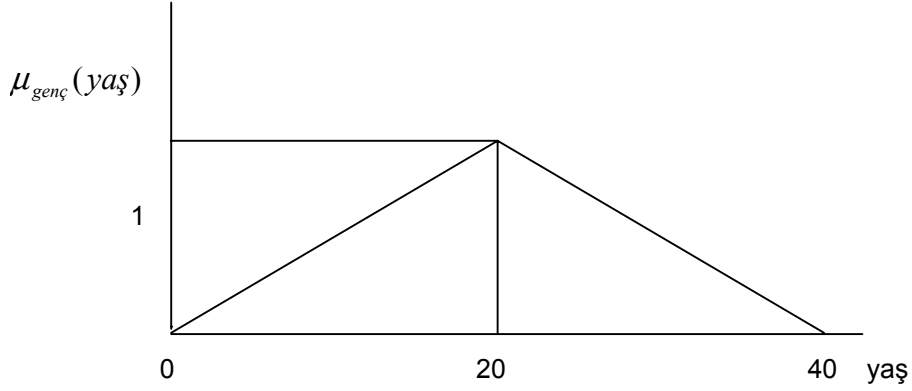
Açık veya net olmayan yada bulanık olan verilerle pazarlama, ekonomi, kalite kontrol, görüntü tanımlama, yapay zeka gibi bir çok alanda karşılaşırız. Bu kesin olarak tanımlanmayan veya sınıflanmayan veriler için bulanık sayılar veya veriler tanımını kullanılmaktadır. Değişkenin aldığı sayısal değer subjective bir tanımlamaya yol açıyorsa bu veri bulanık veri olarak tanımlanır [3,6,9,10].

Yaşlı, genç, kilolu, zayıf gibi tanımlamalar bulunulan ortama veya kişiden kişiye değişir. Ayrıca, yaşı 20 olan bireyi genç olarak tanımlarsak 25 yaşındaki bireyi nasıl tanımlarız. Bulanık veriler tanımlandıkları kümelere ait üyelik fonksiyonları ile tanımlanan üyelik derecelerine sahiptirler. Aşağıdaki grafikte genç nüfusu tanımlayan bir üçgensel üyelik fonksiyonunu ve gözlenen bir bireyin bulanık yaş değerine göre ne kadar bu kümeye ait olduğunu yada ne kadar genç olduğunu tanımlamaktadır. 20 yaşındaki birey 1 olasılıkla bu gençler kümesinin bir elemanıdır. 10 yada 30 yaşındaki bireyler genç tanımına %50 uymaktadır. 20 yaşından uzaklaştıkça buradaki tanımlamaya göre genç tanımlaması yanlış olmaktadır. Bulanık verileri

<sup>1</sup> Gazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü

<sup>2</sup> Selçuk Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü

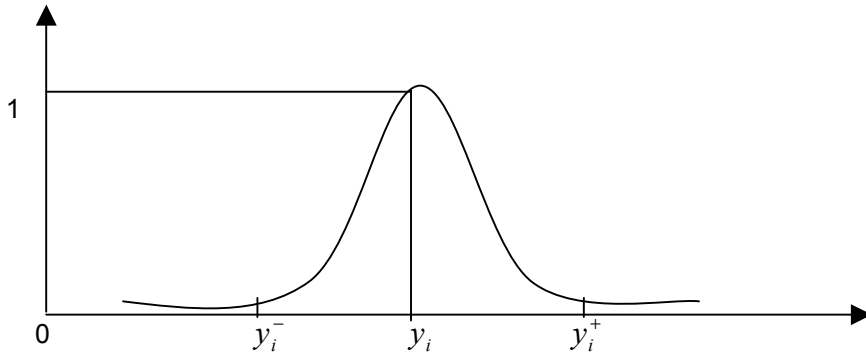
yapısı gereği farklı üyelik fonksiyonları ile tanımlamak gerektiğinden çok farklı üyelik fonksiyonları mevcuttur ve uygulayıcının kendisi de farklı bir bulanık küme fonksiyonu tanımlayabilir.



Şekil 1. Üçgensel gençlik üyelik fonksiyonu

Bilinen klasik regresyon problemlerinde gözlemler sabit yada değişmez olarak kabul edilir. Örneğin, ağırlık ve boy arasındaki varsayılan doğrusal modelin tahmininde her bir bireyden elde edilen gözlem yada ağırlık ve boy değeri sabittir. Belirli bir bireyin ağırlığı için bir boy değeri yada boyu için bir ağırlık değeri gözlenir.

Yakalama-tekrar yakalama yönteminde elde edilen gözlemlerde bağımlı değişken olan yığındaki işaretli oranı ( $Y$ ) bağımsız değişken olan yığındaki işaretli sayısı ( $X$ ) için belirli bir değer alması söz konusu değildir. Belirli bir yığındaki işaretli sayısı ( $x_i$ ) için alınacak  $n_i$  hacimli bir örnekten yığındaki işaretli oranı ( $y_i$ ) örneğe bağlı olduğundan belirli bir değer almaz. Dolayısıyla  $x_i$  değerinde örneğe bağlı olarak farklı  $y_i$  değerleri bulmak doğaldır ve bu durumda  $y_i$  değeri sabit olmayan yada bulanık bir sayı olarak  $i$  nci adımdaki yığındaki işaretli oranını ifade eder. Gerçekte  $y_i$  değeri  $i$  nci adımdaki yığın  $\pi_i$  işaretli oranının nokta tahminidir.  $y_i$  tahminleri örneğe bağlı olduğundan yada doğrudan bir gözlem değeri olmadığından bulanık bir sayıdır. Bu  $y_i$  bulanık sayısı  $\pi_i$  değerini %95 ihtimalle kapsayacak şekilde aşağıdaki üyelik fonksiyonu ile tanımlanabilir.



Burada,

$$y_i^- = y_i - \sqrt{\frac{y_i(1-y_i)}{n_i}} \quad 1,96 \quad \text{ve} \quad y_i^+ = y_i + \sqrt{\frac{y_i(1-y_i)}{n_i}} \quad 1,96 \quad (1)$$

olarak tanımlanır. Gerçekte,  $i$  nci yakalama-tekrar yakalama adımında  $Y_i$  değişkeni beklenen değeri  $\pi_i$  olan binom dağılımına ( $Y_i \in \text{bin}(n_i, \pi_i)$ ) sahiptir. Yakalama-tekrar yakalama metodunda  $x_i$  bağımsız değişkenine karşılık gelen  $y_i$  bağımlı gözlem değeri ( $y_i^-, y_i^+$ ) aralığında değişen bulanık bir sayıdır. Bu durumda, regresyon modelinde ters tahmin yönteminin kullanılması için iki bulanık regresyon modelini inceleyelim[1,2,4,5,8,11].

### Bulanık Regresyon Modelleri

İnceleyeceğimiz iki bulanık regresyon modelini aynı örnek üzerinde uygulayalım. Bu iki model farklı varsayımlar üzerine kurulmuştur. Modellerden birini tercih ederken modellere ilişkin bu varsayımların mutlak incelenmesi gerekir. Veriler 7 günlük süre içerisinde yakalama-tekrar yakalama yöntemi ile elde edilen gerçek bir çalışmadan alınmıştır [7].

#### Model 1

| Gün(i) | $n_i$ | Örnekteki işaretli sayısı ( $r_i$ ) | Yığındaki işaretli sayısı ( $x_i$ ) | Örnekteki işaretli oranı $y_i = (r_i/n_i)$ | $y_i^-$  | $y_i^+$  |
|--------|-------|-------------------------------------|-------------------------------------|--|----------|----------|
| 1      | 437   | 0                                   | 0                                   | 0  | *        | *        |
| 2      | 890   | 128                                 | 437                                 | 0.14382                                    | 0.120766 | 0.166874 |
| 3      | 1019  | 409                                 | 1199                                | 0.40137                                    | 0.371273 | 0.431467 |
| 4      | 88    | 55                                  | 1809                                | 0.62500                                    | 0.523849 | 0.726151 |
| 5      | 83    | 48                                  | 1842                                | 0.57831                                    | 0.472069 | 0.684551 |
| 6      | 45    | 30                                  | 1877                                | 0.66666                                    | 0.528925 | 0.804395 |
| 7      | 77    | 47                                  | 1892                                | 0.61038                                    | 0.501454 | 0.719306 |

Model 1,  $y_i$  bulanık gözlemini ifade eden doğrusal model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , \quad i = 1, 2, \dots, k \quad (2)$$

$$y_i^- = \beta_0^- + \beta_1^- x_i + \varepsilon_i^- \quad , \quad i = 2, \dots, k \quad (3)$$

ve

$$y_i^+ = \beta_0^+ + \beta_1^+ x_i + \varepsilon_i^+ \quad , \quad i = 2, \dots, k \quad (4)$$

doğrusal modelleri arasında kalacaktır. Bu üç modelden yığın hacmi  $N$  için (2) modeli, yığın hacminin alt sınırı  $N_{alt}$  için (4) ve yığın hacminin üst sınırı  $N_{üst}$  için (3) modelinde ters tahmin yöntemi kullanılarak tahminler elde edilir. Tablodaki verileri Model 1'e uygularsak doğrusal modeller şöyle olur:

$$\hat{y}_i = -0,000762 + 0,0003347x_i \quad (2')$$

$$\hat{y}_i^- = 0,02178 + 0,0002637x_i \quad (3')$$

$$\hat{y}_i^+ = -0,02560 + 0,0004070x_i \quad (4')$$

Bu modellerden yığın hacmi, alt sınırı ve üst sınırının tahmini ters tahmin yöntemi ile aşağıda belirlenmiştir. Bu alt ve üst değerler  $z = 1.96$  ile %95 güven sınırlarıdır.

$$\hat{N}_{alt} = \frac{1 - (-0,0256)}{0,000407} \cong 2518$$

$$\hat{N} = \frac{1 - (-0,000762)}{0,0003347} \cong 2990$$

$$\hat{N}_{üst} = \frac{1 - (0,02178)}{0,0002637} \cong 3710$$

Gerçekte binom dağılımından gelen  $y_i$  bulanık gözlemleri normal dağılım bulanık fonksiyonu ile tanımlanmıştır. Burada yapılan en önemli varsayımlar:  $y_i$  gözlem değerleri çok küçük değerler almayacak ve örnek hacimleri oldukça yüksek olacaktır. Bu şartların sağlanması yada varsayımların geçerli olması durumunda binom dağılımını normal dağılıma dönüştürmek ve bulanık bir sayı olarak tanımlamak doğru olacaktır.

## Model 2

$Y_i$  değişkeninin  $n_i$  ve  $\pi_i$  parametrelili binom dağılımına sahip olduğunu biliyoruz.  $i \geq 2$  için  $y_i$  gözlem değeri daha önce örnek yada işaretli sayılarına bağlıdır.  $y_i$  gözleminin beklenen değeri

$$E(y_i) = \frac{x_i}{N} \quad (5)$$

eşitliği ile ifade edilir. Burada  $x_i$ , yakalama-tekrar yakalama yönteminde örneğe alınan birimlerin işaretli sayısını, başka bir ifadeyle yığındaki işaretli birimlerin toplam sayısını ifade eder. Bu sayı daha önce örneğe alınan birimlerin miktarına bağlıdır. Her  $y_i$  gözlemi daha önceki gözlemleri oluşturan örnek hacimlerine yada daha önceki gözlemlerin dağılımına bağlıdır. Çünkü binom dağılımları örnek hacmi  $n$  ve yığın oranı  $\pi$  parametrelerine bağlıdır.

$y_i$  bulanık gözlemlerinin bağımlı olması sebebiyle yine normal bulanık küme fonksiyonu ile bu bulanık sayıların sınırlarını şöyle ifade edebiliriz;

$$y_i^- = \hat{y}_i - \sqrt{\frac{\hat{y}_i(1-\hat{y}_i)}{n_i}} \cdot 1,96 \quad , \quad i = 2, \dots, k \quad (6)$$

$$y_i^+ = \hat{y}_i + \sqrt{\frac{\hat{y}_i(1-\hat{y}_i)}{n_i}} \cdot 1,96 \quad , \quad i = 2, \dots, k \quad (7)$$

Burada

$$\hat{y}_i = b_0 + b_1 x_i, \quad i = 1, 2, \dots, k \quad (8)$$

bağımlı gözlemlerin genel modeli olarak tanımlanır. (6) ve (7)'deki bulanıklık sınırları genel bağımlılık dolayısıyla  $y_i$  gözlemlerine değil genel yapıyı temsil eden  $\hat{y}_i$  değerlerine bağlıdır. Alt ve üst genel bulanıklık sınırları aşağıdaki iki model ile tahmin edilir:

$$\hat{y}_i^- = b_0^- + b_1^- x_i, \quad i = 2, \dots, k \quad (9)$$

$$\hat{y}_i^+ = b_0^+ + b_1^+ x_i, \quad i = 2, \dots, k \quad (10)$$

Bu modeli aynı örneğe uygulayalım.

| G<br>ün(i) | $i$ | Örnekteki<br>işaretli<br>sayısı ( $r_i$ ) | Yığın<br>indaki<br>işaretli<br>sayısı ( $x_i$ ) | Örnekteki<br>işaretli<br>oranı<br>$y_i = (r_i/n_i)$ | $\hat{y}_i$ | $y_i^-$  | $y_i^+$  |
|------------|-----|---|---|---|-------------|----------|----------|
| 1          | 37  | 0   | 0   | 0   | 0.000762    | *        | *        |
| 2          | 90  | 128                                       | 437   | 0.14382   | 0.455019    | 0.122335 | 0.168667 |
| 3          | 019 | 409                                       | 1199  | 0.40137   | 0.005433    | 0.370456 | 0.430629 |
| 4          | 8   | 55  | 1809  | 0.62500   | 0.047103    | 0.502558 | 0.706862 |
| 5          | 3   | 48  | 1842  | 0.57831   | 0.157554    | 0.511108 | 0.720401 |
| 6          | 5   | 30  | 1877  | 0.66666   | 0.274699    | 0.486207 | 0.768732 |
| 7          | 7   | 47  | 1892  | 0.61038   | 0.324904    | 0.524801 | 0.740179 |

Tabloda (8) modelinden  $\hat{y}_i$  değerleri oluşturulmuş ve bu değerlere bağlı olarak (6) ve (7) eşitliklerinden  $y_i^-$  ve  $y_i^+$  değerleri hesaplanmıştır. Bulanık alt ve üst sınırlar için (9) ve (10) modellerini tahmin edelim.

$$\hat{y}_i^- = 0.02304 + 0.0002627x_i, \quad i = 2, \dots, k \quad (9')$$

$$\hat{y}_i^+ = -0.0246 + 0.0004067x_i, \quad i = 2, \dots, k \quad (10')$$

(8), (9) ve (10) modellerinden ters tahmin yöntemi ile yığın hacminin tahmini değişmeyecek (2990) fakat yığın hacmi için limit değerleri biraz farklılık göstererek alt sınır 2519 ve üst sınır 3719 olarak hesaplanmıştır.

**Sonuç**

Bu çalışmada gerekçeleri ile iki farklı model bulanık  $y_i$  gözlemlerine uygulanmış ve sonuçların çok farklılık göstermediği gözlenmiştir. Yığın hacminin tahminin daha doğru olması ve bu modellerin geçerliliğinin sağlanması için yakalama-tekrar yakalama tekniğinin uygulanması sırasında örneğe alınan birimlerin mümkün olduğu kadar çok ve örnekteki işaretli birimlerin oranlarının yüksek olması gereklidir.

### Kaynaklar

1. A. Bardossy, Note on fuzzy regression, Fuzzy Sets and Systems, 37, 65-75, 1997.
2. A. Celmins, Least squares model fitting to fuzzy vector data, Fuzzy Sets and Systems, 22, 245-269, 1987.
3. B. Werners, An interactive fuzzy programming system, Fuzzy Sets and Systems, 23, 131-147, 1987.
4. D. T. Redden and W. H. Woodall, Properties of certain fuzzy linear regression methods, Fuzzy Sets and Systems, 64, 361-375, 1994.
5. G. Peters, Fuzzy linear regression with fuzzy intervals, Fuzzy Sets and Systems, 63, 45-55, 1994.
6. H. J. Zimmerman, Description and optimization of fuzzy systems, International J. General Systems, 4, 1976.
7. H. Mirtaghizadeh, Estimation of population with inverse prediction method using capture-recapture technique, PhD. Thesis, Ankara University, Graduate School of Natural and Applied Sciences, Department of Biometry-Genetic, 2001.
8. H. Tanaka, Fuzzy data analysis by possibilistic linear models, Fuzzy Sets and Systems, 24, 363-375, 1987.
9. H. Tanaka, T. Okuda and J. Asai, On fuzzy mathematical programming, J. Cybernet, 3, 1974.
10. J. L. Verdegay, Fuzzy mathematical programming, in: M. M. Gupta and E. Sanches, Eds., Fuzzy Information and Decision Process, North-Holland, Amsterdam, 1982.
11. W. Xizhao and H. Minghu, Fuzzy linear regression analysis, Fuzzy Sets and Systems, 51, 179-188, 1992.