

Geliş Tarihi : 21.02.2003

## Populasyon Hacminin Yakalama-Tekrar Yakalama Yöntemi Kullanılarak Ters Tahmin Yöntemi ile Tahmini<sup>(1)</sup>

Hamit MİRTAGHIZADEH<sup>(2)</sup>

Tahsin KESİCİ<sup>(3)</sup>

**Özet:** Bu çalışmada, daha önce populasyon hacmi tahmini için önerilen ve en çok kullanılan metotlar sunulmuş ve yeni bir metot olan ters tahmin metodu tanıtılmıştır. Gerçekte bu yöntem, yakalama-tekrar yakalama yöntemine dayalı olarak elde edilen verileri kullanan ve populasyondaki işaretli sayısı ile örnekteki işaretli oranı arasındaki doğrusal ilişkiye dayanan bir metottur. Ankara Üniversitesi, Ziraat Fakültesi'nde yapılan bir uygulama ile diğer metotlar ve ters tahmin metodu mukayese edilmiştir.

**Anahtar kelimeler:** Yakalama-tekrar yakalama, populasyon hacminin tahmini, ters tahmin, regresyon.

### Estimation of Population with Inverse Prediction Method Using Capture-Recapture Technique

**Abstract:** In this study, the methods suggested and mostly used for population abundance estimate were presented and as a new one inverse prediction method was introduced. In fact this is a method that uses the data gathered via capture-recapture process and bases on a linear relationship between the number of marked individuals in population and the ratio of marked individuals in the sample. Other methods and inverse prediction method were compared by an application done at Faculty of Agriculture in Ankara University.

**Key words:** Capture-recapture, estimation of population size, inverse prediction, regression

### Giriş

Bilimsel çalışmalarda üzerinde durulan populasyonun çeşitli özelliklerine ilişkin ortalama ve toplam gibi parametreler, söz konusu populasyondan çekilen örneklerden tahmin yoluyla elde edilir. Gerçekte birçok çalışmada populasyon hacmi bilinmez. Halbuki populasyon hacminin bilinmesi birçok durumda çok önemlidir. Bu sebeple populasyon hacminin tahmini üzerine örnekleme ve tahmin metotları geliştirilmiştir (Cormack, 1968; Cormack, 1972; El-Khorazaty ve ark., 1977; Burnham ve Overton, 1979; Cormack, 1981; Chao, 1986; Cormack, 1989; Darroch ve ark., 1993).

Populasyonların çok geniş ve sayma imkanının bulunmadığı durumlarda, örneklemeyle dayalı yakalama-tekrar yakalama yöntemi ile populasyon hacminin tahmini yoluna gidilmektedir. Herhangi bir bölgede bulunan belirli bir kuş türü populasyonunun, her hangi bir denizde bulunan balık populasyonunun, her hangi bir bölgede belirli bir hastalığı taşıyanların sayısının belirlenmesinde bu yöntemlerden geniş olarak yararlanılmaktadır (Pollak, 1991). Yakalama-tekrar yakalama yönteminin ilk izlerine 17. yüzyılda rastlanmasına karşın bazı teorik bilgiler 1930'larda ortaya çıkmıştır (Lincoln, 1930). Populasyonların kapalı yada değişmezliği de önemli

kavramlardır. Son yıllarda ortaya konulan liste kaynaklarının kullanılması yöntemi Wittes tarafından 1968 yılında önerilmiştir (Wittes ve Sidel, 1968; Wittes, 1970; Wittes, 1974; Wittes ve ark., 1974).

Bu çalışmada, populasyon hacminin tahmini için gerekli bilgileri sağlayacak örnekleme yöntemlerinden bahsedilmiş ve örneklemeden elde edilen bilgileri kullanan tahmin yöntemleri ele alınmıştır. Yakalama-tekrar yakalama yöntemini kullanan tahmin metotları tek yakalama-tekrar yakalama ve seri yakalama-tekrar yakalama metotları olarak tekrar ikiye ayrılmıştır. Regresyon yöntemiyle ters tahmin metodu ve mevcut metotlara ait karşılaştırmalar sunulmuştur.

### Materyal ve Yöntem

#### Materyal

Çalışmanın materyalini, 7 gün boyunca yakalama-tekrar yakalama yöntemiyle örneklenen Ankara Üniversitesi Ziraat Fakültesindeki hazırlık sınıfı hariç 3001 adet 1, 2, 3 ve 4. sınıf öğrencileri oluşturmuştur.

<sup>(1)</sup> Bu çalışma doktora tezinin bir kısmı kullanılarak yazılmıştır.

<sup>(2)</sup> Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Zootehni Bölümü, Biyometri ve Genetik Ana Bilim Dalı ,65080 -VAN

<sup>(3)</sup> Ankara Üniversitesi, Ziraat Fakültesi, Zootehni Bölümü, Biyometri ve Genetik Ana Bilim Dalı, ANKARA

Populasyon hacminin tahmini için, 7 gün boyunca kayıt yaptırmaya gelen öğrencilerin adı, soyadı ve numaraları kaydedilerek 7 liste kaynağı oluşturulmuştur. Öğrencilerden ilk gün gelenlerin hepsi işaretlenmiş (yakalanmış) olarak varsayılmıştır. İkinci gün gelenler içinde ilk günde gelenler işaretli öğrenciler olarak kaydedilmiştir. Bu uygulama 7. güne kadar devam ettirilmiştir. Böylece yakalama-tekrar yakalama yöntemine göre listeler elde edilmiştir. Elde edilen listelerde sadece 1. gün gelenlerin sayısı, sadece 2. gün gelenlerin sayısı ,..., sadece 7. gün gelenlerin sayısı, 1. ve 2. gün gelenlerin sayısı, 1. ve 3. gün gelenlerin sayısı ,..., ve nihayet 7 günün 7' sinde de kayıt yaptırmaya gelen öğrencilerin sayısı şeklinde kayıtlar yapılmıştır.

Kayıt haftasında gelen öğrencilerin isimleri ve öğrenci numaraları alınarak listeler oluşturulmuştur.

**Yöntem**

Ankara Üniversitesi Ziraat Fakültesi öğrencilerinin sayısını tahmin etmek amacıyla liste kaynakları değerlendirilmiştir. Bu şekilde gerçek öğrenci sayısı belirlenmiştir. 4 farklı yakalama-tekrar yakalama yöntemi ile öğrenci sayısı tahminlenmiştir. Kullanılan yöntemler;

- i. Liste kaynaklarını kullanan Genelleştirilmiş Standart Yakalama Tekrar Yakalama Yöntemi (Gmt) (Chao, 1986),
- ii. Sürekli yakalama-tekrar yakalama yöntemlerinden biri olan Schnabel Yöntemi (Schnabel, 1938),
- iii. Sürekli yakalama-tekrar yakalama yöntemlerinden Schumacher-Eschmeyer yöntemi (Pollock, 1991),
- iv. Yakalama tekrar yakalama tekniğini kullanan, Ters Tahmin Yöntemi'dir (Neter ve ark., 1989; Kesici ve Kocabaş, 1998).

Gmt standart yönteminde populasyon hacminin tahmini;

$$\tilde{N} = S + \sum_{i=1}^k \sum_{j=i+1}^k z_i z_j / f_2 \tag{1}$$

şeklindedir. Burada;

- $\tilde{N}$  : Populasyon hacminin tahminini,
- k : liste sayısı ya da ele alınan zaman birimi sayısıdır. Eğer 5 defa yakalama yapılmışsa k=5 tir.
- s : k tane listelerin birleşmesinden kaynaklanan farklı isimlerin sayısı,
- $z_i$  : yalnızca i. listede görülenlerin sayısı,
- $f_2$  : tam olarak listelerde 2 defa yer alanların sayısıdır.

Tahminlenen N sayısının varyansı;

$$\hat{V}(\tilde{N}) \approx (\tilde{N} - S) + (\tilde{N} - S)^2 \left( \frac{1}{f_2} + \frac{4}{\tilde{N}} \right) + \left[ \sum_{i=1}^k (f_1 - z_i)^2 z_i - \left[ \sum_{i=1}^k (f_1 - z_i) z_i \right]^2 / \tilde{N} \right] / f_2^2 \tag{2}$$

olmaktadır. Burada;

$f_1$ : listelerde sadece 1 defa yer alan bireylerin sayısıdır. Gmt yönteminin varsayımları: k listeleri birbirinden bağımsızdır.  $(p_{1j}, p_{2j}, p_{3j}, \dots, p_{Nj})$  ( $j=1, 2, \dots, k$ ) olmak üzere dağılış fonksiyonu  $F_j^j$  den elde edilen bağımsız şansa bağlı örneklerdir  $j=1, 2, \dots, k$  olmak üzere  $(p_{1j}, p_{2j}, \dots, p_{Nj})$  sırasıyla j üzerinden bağımsızdır.

Gmt modelinde populasyon çapı için yürütülen mantık, k listeden elde edilen farklı isimlerin sayısı olan S değerini listelere hiç kaydedilmemiş kişilerin sayısına eklemektedir. Bu durum,

$$N = S + f_0$$

olarak gösterilebilir.  $f_0$  hiçbir listede yer almayan bireylerin sayısı olmak üzere amaç,  $f_0$ 'ın bir tahmin edicisini bulmaktır. Beklenen değeri;

$$E(f_0) = \sum_{i=1}^N E E (\text{listelere hiç kaydedilmemiş bireyler} / p) \tag{3}$$

şeklindedir. Güven aralıkları;

$$\hat{N} \pm Z_{\alpha/2} S(\hat{N}) \tag{4}$$

şeklindedir. Burada  $s(\tilde{N})$  tahminin standart sapmasıdır.

Schnabel Yöntemi, sürekli yakalama-tekrar-yakalama yöntemini kullanan bir tahmin yöntemidir. Populasyon hacminin tahminleyicisi;

$$\hat{N} = \frac{\sum_{i=1}^n N_i M_i}{\sum_{i=1}^n R_i} \tag{5}$$

şeklindedir. Burada;

- $\hat{N}$  = populasyon hacminin tahmini,
- $N_i = t_i$  :  $i$ 'nci örnekte yakalanan bireylerin sayısı,
- $R_i = i$ 'nci örnekte tekrar yakalananların sayısı ( $i = 1, 2, 3, \dots, n$ ),
- $M_i = i$ 'nci örnek alınmadan hemen önce populasyondaki işaretli bireylerin sayısıdır. Dağılışın standart sapması;

$$\sqrt{\hat{V}\left(\frac{1}{\hat{N}}\right)} = \sqrt{\frac{\sum_{i=1}^n R_i}{\left(\sum_{i=1}^n N_i M_i\right)^2}} \tag{6}$$

Schnabel güven aralığı,

$$1/N = \frac{1}{\hat{N}} \pm t_{(\alpha/2, n-1)} \sqrt{\hat{V}\left(\frac{1}{\hat{N}}\right)} \tag{7}$$

şeklindedir. N için güven aralığı ise 7 numaralı eşitliğin tersi alınarak aşağıdaki gibi hesaplanır.

$$\left( \frac{1}{N_{\text{İçinÜst sınır}} / N}, \frac{1}{N_{\text{İçinAlt sınır}} / N} \right) \tag{8}$$

Schumacher-Eschmeyer yöntemi Schnabel modelindeki gibi  $y_i = R_i / N_i$  ve  $x_i = M_i$  grafiğinin doğrusal olması varsayımına dayanmaktadır ve aynı notasyonlar ile aynı varsayımları kabul ederler. Yöntemde populasyon hacminin tahmini;

$$\hat{N} = \frac{\sum_{i=1}^n (N_i M_i^2)}{\sum_{i=1}^n (R_i M_i)} \quad (9)$$

şeklindedir. Varyans;

$$S^2(1/\hat{N}) = \frac{\sum (R_i^2 / N_i) - (\sum R_i M_i)^2 / \sum (N_i M_i^2)}{(n-2)\sum (N_i M_i^2)} \quad (10)$$

olarak elde edilir. Standart hata  $S(1/\hat{N}) = \sqrt{S^2(1/\hat{N})}$  olmaktadır. Güven aralığı;

$$\frac{1}{\hat{N}} \pm t_{\frac{\alpha}{2}; n-1} S(\hat{N}) \quad (11)$$

şeklinde hesaplanmaktadır.

Populasyon hacminin tahmininde kullanılan ve çalışmanın ağırlıklı konusunu teşkil eden Doğrusal Regresyonda Ters Yolla Tahmin yönteminde kullanılan eşitlikler aşağıdaki gibidir.

$$X_i = \sum_{i=1}^{n-1} U_i \quad , \quad (R_1 = 0, X_1 = 0, Y_1 = 0) \quad (12)$$

$N_i = i$ 'nci örnekte yakalanan bireylerin sayısı ( $N_i = R_i + U_i$ ),

$R_i = i$ 'nci örnekte tekrar yakalanan bireylerin sayısı,

$U_i = i$ 'nci örnekte işaretlenen bireylerin sayısı (yada işaretsizlerin sayısı),

$M_i = i$ 'nci örnek alınmadan hemen önce populasyondaki işaretli bireylerin toplam sayısıdır.

$Y_i = i$ 'nci örnekteki işaretli birey oranı;

$$Y_i = \frac{R_i}{N_i} = b_0 + b_1 X(M_i) \quad (13)$$

şeklinde doğrusal olarak açıklanabilir. Populasyon hacminin tahmini;

$$\hat{N} = \frac{1-b_0}{b_1} \quad , \quad (b_1 \neq 0) \quad (14)$$

şeklindedir. Burada;

$$b_0 = \bar{Y} - b_1 \bar{X} \quad (15)$$

$$b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n d_x d_y}{\sum_{i=1}^n d_x^2} \quad (16)$$

şeklinde hesaplanır. Varyans;

$$s^2(\hat{N}) = \frac{HKO}{b_1^2} \left[ 1 + \frac{1}{n} + \frac{(\hat{N} - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right] \quad (17)$$

şeklindedir. Populasyon hacmi için güven aralığı;  $\hat{N}$ , populasyon hacminin tahmini için  $(1-\alpha)$  elde edilmektedir. Güven aralığı;

$$\hat{N} \pm t \left( 1 - \frac{\alpha}{2}; n - 2 \right) s(\hat{N}) \quad (18)$$

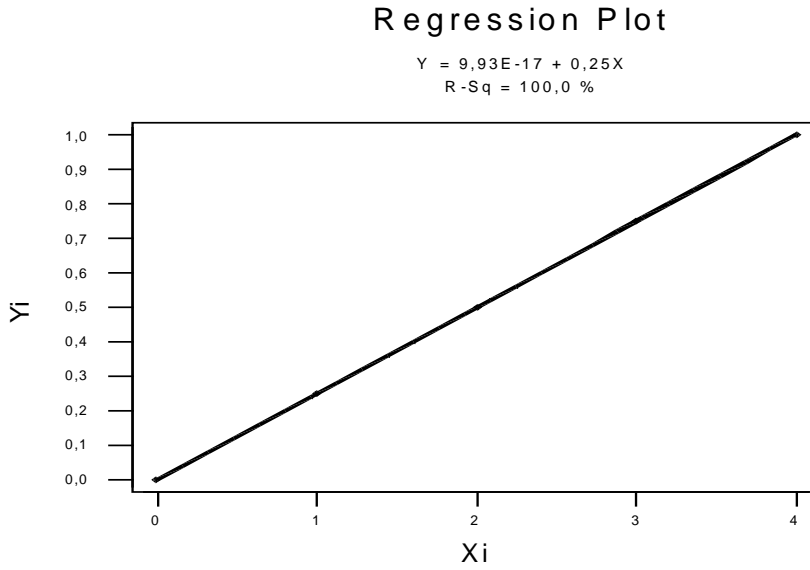
şeklindedir. Geçerli ve tutarlı bir tahmin için iki koşulun sağlanması zorunludur;

$H_0: b_0 = 0$ , hipotezinin red edilmemesi gereklidir.

Şekil 1'deki bilinen doğrusal ilişkiye dayanılarak verilere ait ilişkinin modelinin orijinden geçip geçmediği test edilir.

$H_0: b_1 \neq 0$ , hipotezinin red edilmesi gereklidir (Neter ve ark., 1989).

Ters tahmin yönteminde X ile Y gözlemleri arasında beklenen doğrusal ilişkinin N sayısındaki değişime bağlı olarak grafiksel gösterimi Şekil 1'deki gibidir.



Şekil 1. Ters tahmin yönteminde X ile Y değerleri arasındaki N sayılarına bağlı olarak beklenen değişim.

### Bulgular ve Tartışma

Listelerden yararlanarak, metot bölümünde açıklanan ve kullanılan 4 tahmin metodu kullanılarak elde edilen toplu populasyon hacminin tahmininde yaygın bir şekilde sonuçlar Çizelge 1’de verilmiştir.

Çizelge 1. Yakalama-tekrar yakalama yöntemi ile 7 günde elde edilen veriler

Hafta	Gün	$N_i$	$R_i$	$M_i = X$	$Y_i = \frac{R_i}{N_i}$
1	Pazartesi	437	0	0	0
	Salı	890	128	437	0.143820
	Çarşamba	1019	409	1199	0.401373
	Perşembe	88	55	1809	0.625000
2	Cuma	83	48	1842	0.578313
	Pazartesi	45	30	1877	0.666666
	Salı	77	47	1892	0.610389
Toplam	7	2639			

Çizelge 1’de çalışmada 7 gün boyunca öğrenci kayıtları süresince elde edilen gözlemler özetlenmiştir. Kayıtlar için geçen 7 gün süresince toplam 2639 öğrenci izlenmiştir. Kayıtların ilk gününde gelen öğrenciler arasında daha önce işaretlenen bulunmadığı için  $R=0$ ’dır. Kayıtların ilk gününde gelen tüm öğrenciler listeye alınmıştır. 2. günde ilk gün gelen 437 öğrenci yakalanmış olarak değerlendirilmiş ve bu işlem diğer günlerde de aynı şekilde

devam etmiştir. Kayıtların son gününde toplam 77 öğrencinin 47’si daha önce işaretlenmiş olanlardır.

Elde edilen veri setinden 4 farklı yöntem kullanılarak populasyon hacmi tahminlenmiştir. 4 farklı yöntem için tahminler ve bazı tanıtıcı istatistikler Çizelge 2’deki gibi bulunmuştur.

Çizelge 2. Dört farklı yöntemle populasyon hacminin tahmini, %95'lik güven aralıkları ile varyans ve standart sapmalar

Yöntem	$\tilde{N}$	%95'lik Güven Sınırları		$S^2(\tilde{N})$	$S(\tilde{N})$
		Alt Sınır	Üst Sınır		
Gmt	2998	2820	3176	15587.7	124.851
Schnabel	3002	2796	3240	$1.5 \times 10^{-10}$	$1.244 \times 10^{-4}$
Schmacher-Eschmeyer	3000	2905	3101	$2 \times 10^{-11}$	$4.452 \times 10^{-6}$
Ters Tahmin	2985	2664	3306	8253.0	90.85

Çizelge 2'de dört farklı yöntemle göre populasyon hacmi  $N'$  nin, nokta ve %95 ihtimalli güven aralığı tahminleri özetlenmiştir. Buna göre, gerçek  $N$  değeri olan 3001'e en yakın tahmini Schnabel ve Schmachel-Eschmeyer vermiştir. Bunu 2998 ile Gmt ve 2985 ile Ters-Tahmin yöntemi izlemektedir.

Güven aralıkları incelendiğinde Schmachel-Eschmeyer yönteminin daha dar bir aralıkta populasyon hacmini tahminlemiştir. Ters Tahmin yöntemi ise en geniş tahmin aralığını vermiştir.

Schmacher ve Schmachel-Eschmeyer yöntemleri en küçük varyans ve standart hataya sahiptir. Ancak bu yöntemlerde  $N$  için değil  $1/n$  değeri için hatalar elde edilmektedir.

Populasyon hacminin tahmininde liste kaynaklarını kullanan genelleştirilmiş standart yakalama-tekrar yakalama modeli (Gmt), sürekli yakalama-tekrar yakalama yöntemlerinden Schnabel ve Schumacher-Eschmeyer metotları ve hem liste kaynaklarını hem de yakalama-tekrar yakalama metotlarını kullanan regresyon denkleminde ters yolla populasyon hacmini tahmin eden ve alternatif bir yaklaşım olarak kullanılan bu metot diğerleriyle mukayese edildiğinde, görülmüştür ki doğrusal ilişkiye dayanan tahmin metotlarını kullanmak daha kolay ve pratiktir. Aynı zamanda bu uygulamadan çıkan bir diğer sonuç ise yakalama-tekrar yakalama metotları Gmt kadar duyarlılık göstermişlerdir. Ayrıca Ters Tahmin metodunun sonuçları da oldukça yeterlidir. Ters tahmin metodunun bir avantajı da sapmalı örnekleri regresyon tekniğini kullanarak belirlemesidir. Önerilen ters tahmin yöntemi, diğer yöntemlere nazaran hesaplama adımları oldukça kısa ve kolay, yöntemin gerektirdiği varsayımlar ise daha sınırlıdır. Bunlara ilaveten ters tahmin yöntemi regresyon eşitliğinden yararlanarak bulunduğu için bu yöntem araştırmacılar tarafından daha kolay algılanabilir ve daha yaygın kullanılabilir.

Tahmin sonuçları açısından bu 4 metot karşılaştırıldığında ise gerçek populasyon hacmi olan  $N = 3001$ , 4 yöntemde de çok yakın sonuçlarla tahmin edilmiştir. Ayrıca 4 yöntemle göre tahmin edilen % 95 ihtimalli güven sınırları içerisinde gerçek populasyon hacmi girmektedir.

## Kaynaklar

- Burnham, K. P., W. S. Overton, 1979. Robust estimation of population size when capture probabilities vary among animals. *Ecology*, 60: 927-936.
- Chao, A., 1986. Estimating population size in a generalized capture-recapture Model with Applications to Epidemiological Data. *Statistical Theory and Data Analysis*, 2: 29-36.
- Cormack, R.M., 1968. The Statistics of capture- recapture. oceanography and marine biology. *Annual Review*, 6: 455-506.
- Cormack, R.M., 1972. The logic of capture-recapture estimates. *Biometrics*, 28: 337-343.
- Cormack, R.M., 1989. Interval estimation for mark-recapture-recapture experiments. *Biometrics*, 45: 427-438.
- Cormack, R.M., 1981. *Log-linear Models for Capture-Recapture Experiments on Open Populations*. In: *Hiorns RW, Cooke D, eds. The Mathematical Theory of the Dynamics of Biological Populations II*. London: Academic Press.
- Darroch, J. N., S.E. Fienberg, G.F.G., Glonek, 1993. A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *J. Ame. Stat. Assoc.*, 88: 1137-1148.
- El-Khorazaty, M.N., P.B. Imsey, G.G. Koch, H.B., Welles, 1977. A review of methodological strategies for estimating the total number of events with data from multiple-record systems. *International Statistical Review*, 45: 129-157.
- Kesici, T., Z. Kocabaş, 1998. *Biyostatistik*. Ankara Üniversitesi, Eczacılık Fakültesi yayın no:79, Ankara.
- Lincoln, F.C., 1930. Calculating waterfowl abundance on the basis of banding returns. Circular of the U.S. Department of Agriculture, 118, 1-4.
- Neter, J., W. Wasserman, M.H. Kutner, 1989. *Applied Linear Regression Models*. 2<sup>nd</sup> edition, Richard D. Irwin, Inc.
- Pollock, K.H., 1991. Modeling capture, recapture and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present and future. *J. Ame. Stat. Assoc*, 86: 225-238.

- Schnabel, Z.E., 1938. The estimation of the total fish population of a lake. *American Mathematical Monthly*, 39: 348-352.
- Wittes, J.T., 1970. *Estimation of Population Size: The Bernoulli Census*. Ph.D. thesis, Harvard University, Cambridge, Massachusetts.
- Wittes, J.T., 1974. Applications of a multinomial capture-recapture method to epidemiological data. *J. Ame. Stat. Assoc.*, 69: 93-97.
- Wittes, J.T., T. Colton, V.W. Sidel, 1974. Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *J. Chr. Dis.*, 27: 25-36.
- Wittes, J.T., V.W. Sidel, 1968. A generalization of the simple capture-recapture model with applications to epidemiological research. *J. Chr. Dis.*, 21: 287-301.