ANA KÜTLE TOPLAMININ TAHMİNİNDE ORTALAMA - TOPLAM METODU YERİNE ORAN - TOPLAM METODUNUN KULLANİLMASI ve BİR TARTIŞMA

Doç. Dr. M. Kemal YOĞURTÇUGİL

I.

GİRİŞ

Aynı şartlar altında daima aynı tezahürleri gösteren tipik olaylardan müteşekkil bir topluluğu cüzlerinden biri ile ifade etmek mümkünken, ana vasıflar bakımından bir tek topluluk teşkil eden fakat tâli vasıfları itibariyle ferdi görünüşleri birbirinden farklı olan olayları araştırmak için istatistik tekniği kullanılır. Kısacası bu teknik geniş çaptaki sayısal verilerin değerlendirilmesinde tek vasıtadır ve belli bir veri yığınına uygulanan istatistik teknikler, bu verilerin irdelenmesini mümkün kılarlar.

İstatistik metodlarm gayesi üzerinde çalışılan ana kütle hakkında bir fikir edinmek ve onun vasıflarını tanımaktır. Bu vasıfları karakterize eden rakkamlara parametre denir. Ancak ana kütle içindeki bütün birimleri ele geçirip incelemek imkânsız ve çoğu kere lüzumsuz olduğundan ona ait parametreler bu topluluktan alman bir örnekten hareketle tahmin edilirler. Üzerinde çalışılan numuneyi karakterize eden bu kıymetlere örnek değeri veya statistik denir.

İstatistik araştırmalarında problemin çözümü iki safhalıdır.

a. Kütleye ait parametrelerin yerini tutacak örnek değerlerinin hesaplanması;

b. Bu değerler esas alınarak yığma ait gerçek değerlerin tahmini ve bu tahmindeki geçerlilik sınırlarının diğer bir ifade ile genelleme ile yapılabilecek hatanın derecesinin belirtilmesi.

Başarılı bir örnekleme herşeyden önce:

1. Örnek çekilecek ana kütle hakkında bazı özel bilgilere ihtiyaç gösterir.

2. Seçme işlemi ilgilendiğimiz özellik veya değişkenden bağımsız olmalıdır.

3. Örnek bir ön yargıya yer vermeden ve sistematik bir farklılık yaratmayacak şekilde çekilmelidir.

4. Örneğe alınan birimlerden herbiri yekdiğerinden bağımsız bulunmalıdır.

5. Örneğe alınacak verilerin hepsine aynı şartlar istisnasız uygulanmalıdır.

Verilerin elde edilişinde tek doğru yolun bir tam sayım olduğunda israr edenlerin çoğu, esas dökümanlarda bir çok hata kaynakları bulunabileceği ve yüzde yüzlük bir sayımın bile çok hatalı olabileceği hakikatini ihmal ederler. Gerçekten bugün az sayıda birimlerin tetkikinde hata kaynaklarının çok daha tesirli bir şekilde kontrol edilebilmesi nedeniyle örneklemenin, tatbik edilecek bir tam sayıma göre daha doğru neticeler verdiği kabul edilmiştir. Ancak örnek üzerinden tahminin yüzde yüz sayım ile elde edilmiş parametre ile tamamen aynı olması çok az ihtimal dahilindedir. Bununla beraber bazı örneklerde numune üzerinde tahmin için arzu edilen hassasiyet derecesi daha önceden şart koşulabilir. Bu gibi örneklere ihtimali örnekler denir ve bunların özellikleri her bir birimin seçilme ihtimalinin bilinmesidir. Belli bir ihtimal örneğinin hassasiyetinin bizzat örnekten hareketle tahmin edilebileceği özelliği, bu gibi örnekleri diğer ihtimali olmayan örnekler yerine kullanmak için kuvvetli bir sebep teşkil eder.

Şans ve ihtimal kanunlarına göre geliştirilen istatistik metodları tesadüfi örneklere tatbik edilirler. Bu sebepten istatistik metodları ile elde edilecek neticelerin doğrulukları büyük ölçüde tesadüfi numune almadaki maharete bağlıdır.

II.

TESADÜFİ ÖRNEKLEME

Tamamen aynı olmasına itina edilen şartlar altında tekrar edilen işlemlerde değişik değerler alan değişken'e tesadüfi değişken denir. Bir tesadüfi değişkeni belirlemek onun sayısal değerini tayin etmek değildir; aksi halde tesadüfi değişken olmaktan çıkar. Bu sonuçlar bir $x_i = \{x_1, x_2, \ldots x_n\}$ lojik imkânlar cümlesi teşkil ederler. Bu $\{x_i\}$ cümlesine bir ihtimal ölçüsü verebilmek için $\{x_i\}$'nin her elemanının farklı ve mümkün olan değerlerinin ihtimalleri de verilmiş olmalıdır. Bu ihtimallerin bilinmiş olması halinde bize uygun olan veya daha az uygun olan veyahut hiç uygun olmayan neticelerle ne kadar çok veya ne kadar az karşılaşacağımızı anlarız. Böyle bir cetvelin verilmesi bu tesadüfi değişkenin dağılma kanunun ifade edilmesi demektir; dağılma kanunun bilinmesi ise o tesadüfi değişkene ait tüm soruların cevaplandırılmasını mümkün kılar.

$\{x_1, x_2, \ldots x_n\}$ cümlesi birbirinden bağımsız ve herbiri aynı ihtimal kanununa tâbi n değişkenden meydana gelmişse, bu cümlenin birlikte ihtimal yoğunluğu $f(x_1, x_2, \ldots x_n) = \phi(x_1). \phi(x_2). \phi(x_3). \ldots \phi(x_n)$ dir. Bu takdirde $\{x_i\}$ cümlesine ihtimal yoğunluğu $\phi(x)$ olan bir ana kütleden çekilmiş n birimli bir tesadüfi örnek adı verilir. Bu şekilde seçilmiş n birimli örnekler ortalamalarının ortalaması ana kütle ortalamasına ve varıyansı ise ana kütle variyansının n de birine eşittir. Bu sonuçlar ana kütle bölünmesi nasıl olursa olsun geçerlidir ve bilinen şudur ki n büyüdükçe bu bölünme normale yaklaşır.

Eğer ana kütle bölünmesi normal ise:

a. $\bar{x}$ ve $s^2[E(s^2) = S^2]$ bağımsız şekilde dağılırlar.

b. $\bar{x}$; $N(\bar{X}; \frac{S}{\sqrt{n}})$ olan bir normal bölünme gösterir.

c. s^2; $X^2_{n-1} \cdot \frac{S^2}{n}$ şeklinde bir bölünme verir.

Ana kütle bölünmesinin normal olduğu bilinmese bile, bu yığının ortalaması $\bar{X}$ ve variyansı S^2 ise, bu topluluktan çekilen n birimli, tesadüfi örnekler ortalamaları ortalaması $\bar{\bar{X}}$ ve variyansı S^2/n olan bir normal bölünmeye yaklaşır ve bu yaklaşım n büyüdükçe artar.

Bir tesadüfi değişken x; muhtemel değerler $x_1, x_2, \ldots . x_n$ leri sırasıyle $f(x_1), f(x_2) \ldots . f(x_n)$ ihtimalleri ile alıyorsa bu takdirde x'in matematik ümidi

$$E(x) = x_1 . f(x_1) + x_2 . f(x_2) + \ldots . + x_n . f(x_n)$$

$$= \sum_1^n x_i . f(x_i) = \bar{X} \text{ dür}^{1}$$

Benzer şekilde

$$E(x - \bar{X})^2 = \sum_1^n (x_i - \bar{X})^2 . f(x_i) = \text{var}(x)$$

yazılır.

$$E(x_1 + x_2 + \ldots . + x_n) = E(x_1) + E(x_2) + \ldots . + E(x_n)$$

$$\text{var}(x_1 + x_2 + \ldots . + x_n) = \sum_1^n \text{var}(x_i) + 2 \sum_1^n \text{kov}(x_i, x_j)$$

ve eğer tesadüfi değişkenler yekdiğerinden bağımsız ise

$$\text{var}(x_1 + x_2 + \ldots . + x_n) = \sum \text{var}(x_i) \quad \text{dir.}$$

Şimdi ihtimal bölünmesi $\emptyset(x)$ olan bir topluluktan n sayıda $(x_1, x_2, \ldots . x_n)$ bir tesadüfi örnek alınmış olsa $x = x_1 + x_2 + \ldots . + x_n$'in ihtimal bölünmesi bilinmeden x'in ortalama ve variyansı bulunabilir. Gerçekten ihtimali sondaj tarifine göre $x_1, x_2 \ldots . x_n$, aynı $\emptyset(x)$ ihtimal bölünmesine sahip olduklarından

$$E(x_1) = E(x_2) = \ldots . = E(x_n) = E(x) \quad \text{ve}$$

$$\text{var}(x_1) = \text{var}(x_2) = \ldots . = \text{var}(x_n) = \text{var}(x)$$

den hareketle

$$E(x_1 + x_2 + \ldots . + x_n) = n.E(x)$$

$$\text{var}(x_1 + x_2 + \ldots . x_n) = n.\text{var}(x) \quad \text{dir.}$$

---

1) Eğer x sürekli değişken ise entegral sonlu bir değer olmak şartıyle $E(x) = \int_{-\infty}^{\infty} x.\emptyset(x).dx$ yazılır.

(x) yerine n ile bölünmüş şekli $(\bar{x})$ örnekleme teorisinde çok daha fazla kullanılmakta olduğundan

$$\bar{x} = \frac{1}{n} x_1 + \frac{1}{n} x_2 + \ldots + \frac{1}{n} x_n$$

$$E(\bar{x}) = \frac{1}{n} E(x_1) + \frac{1}{n} E(x_2) + \ldots + \frac{1}{n} E(x_n) \quad \text{veya}$$

$$E(\bar{x}) = \frac{1}{n} n.E(x) = E(x) = \bar{X} \quad \text{ve}$$

$$\text{var}(\bar{x}) = \frac{1}{n^2} \text{var}(x_1) + \frac{1}{n^2} \text{var}(x_2) + \ldots + \frac{1}{n^2} \text{var}(x_n)$$

$$\text{var}(\bar{x}) = \frac{1}{n^2} n.\text{var}(x) = \frac{1}{n} \text{var}(x) \quad \text{dir.}$$

III.

ORTALAMADAN HAREKETLE ANA KÜTLE TOPLAMININ TAHMİNİ veya ORTALAMA - TOPLAM METODU

3.1. Basit Tesadüfi Örnekleme :

Ortalamadan hareketle ana kütle toplamının tahmini ana kütle mevcudu N'in bilinip bilinmediğine göre farklı şekillerde yapılabilir.

A. Ana kütle mevcudu biliniyorsa

Toplamın tahmini $\hat{X} = N.\bar{x}$ ve

bu tahminin variyansı

$$\text{Var}(\hat{X}) = \text{Var}(N.\bar{x}) = N^2. \text{Var}(\bar{x}) = N^2. \frac{S^2}{n}$$

olacaktır.

Bu değer sondaj nisbeti $f = n/N$'in 0.05'den küçük olduğu hallerde veya sonlu ana kütleden çekimin iadesiz yapıldığı durumlarda

$$\mathrm{Var}\,(\hat{X}) = N^2 \cdot \frac{S^2}{n} \cdot \frac{N-n}{N}$$

ile bulunmalı, ana kütle variyansı S^2 meçhul bir değer ise

$$s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1}$$

ile hesaplanan örnek değeri bu kıymetin bir tahmini olduğundan

$$\mathrm{Var}\,(\hat{X}) = N^2 \cdot \frac{s^2}{n} \quad \text{veya}$$

$$= N^2 \cdot \frac{s^2}{n} \cdot \frac{N-n}{N}$$

kullanılmalıdır.

B. Ana kütle mevcudu bilinmiyor ise

a. Sabit bir n değeri için sondaj nisbeti n/N'in veri olduğu hallerde

$$\hat{X} = (N/n) \cdot x = F \cdot x$$

alınmalıdır. Bu tahminin hatası ise x'in standart hatasının F katına eşittir.

$$\mathrm{Var}\,(x) = \mathrm{Var}\,(n\,\bar{x}) = n \cdot S^2$$

$$\mathrm{Var}\,(F \cdot x) = F^2 \cdot \mathrm{Var}\,(x) = N^2 \cdot \frac{S^2}{n}$$

Bu tahminin, n'in sabit olduğu Basit Tesadüfi Sondaj metodları için $N.\bar{x}$'e eşdeğer olduğu; f'nin sabit fakat n'in değiştiği örnekleme modellerinde $F.x$ ile $N.\bar{x}$'in önemli derecede fark ettiği söylenebilir. Çünkü x örnek toplamı olduğu için müşahede hatalarından doğan bir taraflılığa sahiptir; halbuki $\bar{x}$ bir ortalama olduğu için bu taraflılık kaybolacaktır.

b. N bilinmiyor fakat başka kaynaklardan hareketle iyi bir tahmini yapılabiliyorsa $(\hat{N})$, bu takdirde yine $\hat{N}.\bar{x}$ kullanılmalıdır. İki tesadüfi değişkenin variyansı

$$\mathrm{Var}(\hat{N}.\bar{x}) = \hat{N}^2 \cdot \mathrm{Var}(\bar{x}) + \bar{x}^2 \cdot \mathrm{Var}(\hat{N}) + 2\hat{N}\bar{x}\,\mathrm{Kov}(\hat{N}, \bar{x}) \quad \text{dir.}$$

Bu eşitlik genellikle örnek ortalaması ile ana kütle hacmı arasında çok sıkı bir bağlılık olamıyacağından

$$\mathrm{Var}(\hat{N} . \bar{x}) = \hat{N}^2 \mathrm{Var}(\bar{x}) + \bar{x}^2 \mathrm{Var}(\hat{N})$$

haline gelecektir.

$$\frac{\mathrm{Var}(\hat{N})}{\hat{N}^2} = d^2 . \frac{\mathrm{Var}(\bar{x})}{\bar{x}^2}$$

ile gösterirsek

$$\mathrm{Var}(\hat{N}.\bar{x}) = \hat{N}^2 \mathrm{Var}(\bar{x}) (1+d^2)$$

elde edilir. $\hat{N}$ belli ise $\mathrm{Var}(\hat{N}) = d^2 = 0$ olur ve eşitlik

$$\mathrm{Var}(\hat{N}.\bar{x}) = \hat{N}^2 \mathrm{Var}(\bar{x})$$

şekline dönüşür. $\hat{N}$ değişiyor fakat $d = 0.1$ civarında kalıyorsa $\hat{N} = N$ alabiliriz.

(p) bir binom bölünme değişkeni ise

$$\hat{N}.\bar{x} = \hat{N}.p \qquad \text{ve}$$

$$\mathrm{Var}(\hat{N}.\bar{x}) = \hat{N}^2 . p^2 \frac{1-p}{np} + \frac{\mathrm{Var}(\hat{N})}{N^2}$$

değerleri hesaplanmalıdır. (p) bire yaklaştıkça birinci terim ikinciye nazaran küçülecek, (p) sıfıra yaklaştıkça büyüyecektir.

C. Ana kütle mütehavvilliğini göz önüne alan zümrelere göre sondaj usulüne girmeden Basit Tesadüfi Sondajda tahminlerin isabetini arttırmak için bir diğer yolda bu olaya ait iki vasfın şıkları arasındaki münasebetten faydalanarak en küçük kareler metodu yardımıyla tayin edilen bir regresyon doğrusunu tahminde kullanmaktır.

Bu usulde ana kütle ortalamasının tahmini

$$\hat{\bar{X}} = \bar{x} + b\,(\bar{Y} - \bar{y}) \qquad \text{dir.}$$

Bu tahminin variyansı ise

$$\mathrm{Var}(\hat{\bar{X}}) = \frac{N-n}{n.N} S^2 (1 - R^2) \quad \text{olacaktır.}$$

Ancak gerek S^2 ve gerek R^2 ana kütle değerleri olduğundan bunlar yerine tahminleri olan numune değerlerini koymak suretiyle

$$\mathrm{var}(\hat{\bar{X}}) = \frac{N-n}{n.N} s^2 (1 - r^2)$$

elde edilir. Bu usul ile yapılan tahminin daima bir sistematik hata ihtiva ettiği, buna karşılık tahminlerin variyansının doğrudan doğruya tahmin usulünden daha küçük çıktığı kabul edilmektedir.

Pratik ihtiyaçlar bakımından, aynı olay hakkında yapılması mümkün bütün tahminler ortalamasının gerçek değere uygun olup olmadığı hususundan çok münferit bir tahminin bu değere ne kadar yakın olduğu konusu önemli olduğundan, bu kusuruna rağmen Basit Tesadüfi Sondaj'da regresyon usulü tercih edilmektedir.

3.2. Zümrelere Göre Örnekleme :

Herhangi bir örnekleme plânı başlıca iki değişkeni kapsar :

a. Tahminlerin variyansı ve

b. Örneklemenin maliyeti.

$\sigma^2_{\bar{x}} = \dfrac{S^2}{n}$'de $\sigma^2_{\bar{x}}$'in küçültülebilmesi, S^2 veri olduğu için n'in artışlarma bağlıdır. Bu tür örnekleme genellikle

—. Çerçevenin eksikliği, hazırlanması veya tamamlanmasının büyük masrafları gerektirmesinden ve

—. Örneğe dahil birimlerin yekdiğerinden uzak yerleşmesinden doğan büyük müşahede masraflarını da birlikte getirir. Örnekleme maliyetinin veri olduğu durumlarda bu sebepten n gelişigüzel arttırılamamakta ve neticede bir sabiti haline gelmektedir. Problem S^2 ve n'in veri olduğu bir örnekleme plânında $\sigma^2_{\bar{x}}$'nin hangi şekilde küçülebileceğidir.

Optimum usule göre zümrelere göre sondaj'da variyansı büyük olan zümreler için sondaj nisbeti variyansı küçük olanların aksine arttırılmaktadır. Böylece dağıtım zümre standart sapmalarına göre yapıldığından ya sabit mevcutlu örnekler için minimum variyansı veya sabit bir variyansa göre en küçük örnek mevcudunu vermektedir.

Toplamın tahmini

$$\hat{X} = \Sigma N_h . \bar{x}_h \qquad \text{ve}$$

bu tahminin variyansı

$$\mathrm{Var}(\bar{x}) = \frac{1}{n} \left(\frac{\Sigma N_h S_h}{N} \right)^2 \quad \text{den}$$

$$\mathrm{Var}(\hat{X}) = N^2 . \mathrm{Var}(\bar{x}) = \frac{1}{n} (\Sigma N_h S_h)^2$$

değerine eşittir.

Zümre mevcutları N_h'lar belli değilse bu tahmin

$$\hat{X} = \Sigma \frac{n_h . N}{n} . \bar{x}_h$$

la yapılmalıdır.

Zümrelere göre sondaj'da basit tesadüfi sondaj'a nazaran bir kazanç elde edebilme, zümre ortalamalarının birbirinden çok farklı çıkacak şekilde homojen zümreler kurabilme imkânına bağlı olduğu unutulmamalıdır. O halde zümre ortalamaları tahmin edilmek istenen ortalama etrafında toplanıyorlarsa bu takdirde zümrelere gitmeden sadece örnek mevcudunda küçük bir artış yapmak suretiyle aynı kazancı elde etmek mümkündür.

Yüksek müşahede masrafı belirtilen sebeplerle zümrelere göre sondaj için de geçerli olduğu ve tatbikatta daha çok örnekleme maliyeti önem kazandığı göz önünde tutularak aşağıda düşük müşahede masraflı örnekleme plânları «kümelere göre sondaj ve kademeli sondaj» verilmeye çalışılacaktır.

3.3. Kümelere Göre Örnekleme:

Birimlerin müşahedesinin zor ve pahalı olduğu araştırmalarda örnek birimi olarak ana kütle birimlerinin birkaçının bir araya getirilip küme adının verildiği topluluklar kullanılır. Bir şehirdeki binalar yerine binalardan ibaret blokların alınması gibi. Böylece müşahede masrafı düşürülmekte, düşük müşahede masrafı toplam maliyet içinde örnek hacmının artmasına imkân sağlamakta ve muayyen bir para için n'in artması örnekleme variyansını düşürebilmektedir.

Herbiri B birim ihtiva eden A kümeden ibaret bir ana kütleden a küme tesadüfen çekilmiş olsun.

Ana kütle mevcudu $N = A.B$

Örnek mevcudu $n = a.B$

ve ana kütleye dahil her birimin çekilecek örneğe girme ihtimali

$$\frac{a}{A} = \frac{a}{A} \cdot \frac{B}{B} = \frac{n}{N} = f$$

dir. n birimli örnek ortalaması yine ana kütle ortalamasının tarafsız bir tahminidir ve bu aynı zamanda a tane küme ortalamasının ortalamasıdır.

$$\bar{x} = \frac{1}{a} (\bar{x} + \bar{x}_2 + \ldots + \bar{x}_a)$$

$$\mathrm{Var}(\bar{x}) = (1\text{-}f) \frac{S_a^2}{a}$$

$$S_a^2 = \frac{1}{a-1} \sum (\bar{x}_a - \bar{x})^2$$

Basit Tesadüfi Sondaj formülünü andıran bu formülde de ortalamanın variyansı örnek birimleri arasındaki variyansla doğru, örnek birimleri ile ters orantılıdır. $(1-f)$ tabiatıyle iadeli usulde ve küçük f ler için ihmal edilecektir.

Toplamın tahmininin variyansı ise

$$\mathrm{Var}(N . \bar{x}) = N^2 . \mathrm{Var}(\bar{x}) = N^2 (1-f) \frac{S_a^2}{a} \quad \text{dır.}$$

Aynı birimli iki örnekten kümelere göre olanı daha büyük variyans fakat daha düşük maliyet verir. Bu tür örneklemede maksimum etkinlik

a. Aynı kümeye giren birimler arasında farkların büyük

b. Kümeler arası farkların küçük

olmasına diğer bir ifade ile, ortalamaları birbirine yakın kümeler teşkil edebilme imkânına bağlıdır. Örneğe girecek birimlerin ana kütle içinde dağılmış olması tahminde isabeti arttırır ama bu aynı ölçüde maliyeti de fazlalaştırır. Bunun için bir çare kademeli sondaj uygulamasıdır.

Bu örnekleme evvelâ kümelerden bir örnek sonra bunların içinden tekrar bir örnek alma işlemine dayanır. Diğer bir ifade ile önce A tane kümeden a tane küme çekiyoruz. Bu a tane örneğin herbiri B birim ihtiva ediyor. Bundan da b tane alıyoruz.

$$f = \frac{n}{N} = \frac{a}{A} \cdot \frac{b}{B} = f_a \cdot f_b$$

Bu şekilde elde edilen örnek ortalaması da ana kütle ortalamasının tarafsız bir tahminidir.

Ana kütle toplamının tahmini ise

$$\hat{X} = \frac{A}{a} \sum \frac{B}{b} x \qquad \text{veya}$$

$$b = f \cdot \frac{A}{a} \cdot B$$

sayıda tesbit edilmişse

$$\hat{X} = \frac{1}{f} \sum x_i \qquad \text{olur.}$$

Kademeli sondaj'da ana kütle toplamının tahminin variyansı iki kısımdan ibarettir.

$$\mathrm{Var}(\hat{X}) = \frac{A^2}{a} \cdot \frac{A-a}{A} \cdot S_1^2 + \frac{A}{a} \sum \frac{B^2}{b} (1-f)\, S_2^2$$

$$S_1^2 = \frac{1}{a-1} \sum (\bar{x}_a - \bar{x})^2 \qquad S_2^2 = \frac{1}{b-1} \sum (x_i - \bar{x})^2$$

Burada variyansa tesir eden kümeler arası variyansı gösteren birinci terimdir. Bu sebepten :

a. Küme büyüklükleri eşit tutulursa S_1^2 küçülür.

b. Kümeler homojen ise S_2^2 küçülür.

c. A sabit a'yı arttırırsak toplamın variyansı küçülür.

d. $A=a$ alınırsa bu bir zümrelere göre sondaj uygulaması olacak, toplamın birinci terimi sıfır ve neticede variyans sadece ikinci terime bağlanacaktır.

$$\text{Var}(\hat{X}) = \sum \frac{B^2}{b} (1-f)\, S^2$$

Bu ise

$$\text{Var}(\hat{X}) = \sum \frac{N_h^2}{n_h} (1-f)\, S_h^2$$

ile beliren zümrelere göre sondaj'da toplamın tahminin variyansını veren formülden başka bir şey değildir. Zümrelere göre sondaj'da örnek bütün zümrelerden alındığı halde kümelere göre sondaj'da örnek bütün kümelerden alınmıyor. Bu sebepten kümeler arası variyans sıfır olamaz. Şunu da ilâve edelim ki, kümelere göre sondaj'da S_1 mümkün olduğu kadar küçük çıkacak şekilde bir örnekleme plânı tanzimi esas iken, zümrelere göre sondajda S_2^2'yi azaltma düşünülür.

Zümrelere göre sondaj ile kümelere göre kademeli sondaj birlikte uygulanabilir. Bu tür örnekleme plânı ana kütlenin evvelâ zümre kriterine göre zümrelere ayrılması ile başlar; daha sonra her zümre içindeki yığın ana kütle kabul edilerek kümelere bölünür ve bu kümelere yukarıda belirtilen şekilde kademeli sondaj uygulanır. Sonuçta

$$\hat{X}_i = \frac{A}{a} \sum \frac{B}{b} x$$

ile beliren zümre toplamlarının toplamı ana kütle toplamını verecek, bu tahminin variyansı her zümre için

$$\sigma_h^2 = \frac{A_h}{a_h} \left(\frac{A_h - a_h}{A_h}\right) S_{1h}^2 + \frac{A_h}{a_h} \sum \frac{B_h}{b_h} (1-f_h)\, S_{2h}^2$$

ile hesaplanan bağımsız variyansların toplamı olacaktır.

N birimli ana kütle içinde muayyen bir (B) vasfını haiz olanların sayısı (X)'i tahmin etmek istediğimizi farzedelim. Bu takdirde uygulanacak bir kademeli örnekleme plânı içinde önce N birimli yığından n birimli bir tesadüfi örnek çekilecektir. Bu örnek içinde başka bir (A) vasfına sahip olanların sayısı m olmuş olsun. Bu m birimli tâli örneğin içinde (B) vasfını gösterenlerin sayısı x ise $\bar{x} = x/m$ ve

$\bar{m} = m/n$ yazılabilir.

$$\mathrm{Var}(\bar{x}) = (1 - f)\, \frac{S_m^2}{m} \quad \text{ve}$$

M; (A) vasfına ana kütle içinde sahip olanların sayısı ise

$$\hat{X} = M \,.\, \bar{x} \qquad \text{dir.}$$

Bu tahminin variyansı

$$\mathrm{Var}(\hat{X}) = \mathrm{Var}(M \,.\, \bar{x}) = (1 - f)\, M^2 \frac{S_m^2}{m}$$

$$= (1 - f)\, (\frac{M}{m})^2 \,.\, m \,.\, S_m^2 \quad \text{dir.}$$

Genellikle M meçhuldur ve bu nedenle $M \,.\, \bar{x}$ yerine tahmin

$$X = N \,.\, \bar{m} \,.\, \bar{x}$$

ile yapılır. Bu duruda tahminin variyansı

$$\mathrm{Var}(\hat{X}) = (1 - f)\, \frac{N^2}{n - 1} \,.\, \bar{m}^2 \,.\, \left[\, \mathrm{Var}\,(x) + (1 - \bar{m})\, \bar{x}^2 \,\right]$$

dır. Yukarıdaki formülde

$$\mathrm{Var}(x) = \frac{1}{m} \left[\, \Sigma\, x_i^2 - \frac{x^2}{m} \,\right]$$

olduğu hatırlanmalıdır.

IV.

ORAN - TOPLAM METODU

Buraya kadar ki tahlillerimizde gördük ki bir ortalamayı ve onun örnekleme variyansını hesapladığımızda süratle ana kütle toplamını tahmin edebiliyor ve bu tahminin variyansını bulabiliyoruz. Bu usule «Ortalama-Toplam» metodu denilmektedir. Bu konuda bir başka hesap şekli daha vardır ki bu metod «Oran-Toplam» metodu ile anılmakta ve usulün esası pay ve paydasının tahmini bir değer olduğu oranın tahmini değerinin bulunmasına dayanmaktadır. Genellikle «Muayyen bir zaman devresinde belirli bir piyasada belirli bir mal veya hizmetten satılabilecek miktarların tayinine yarayan Piyasa Potansiyeli araştırması veya kısaca Pazar Analizi» adlı pazarlama araştırmasında büyük bir kullanılış sahası bulan bu metod aşağıda izah edilmeye çalışılacaktır.

Ekonomik ve teknik bir cüzütam olan işletmeyi, bir mal veya hizmet üretimi gayesiyle biraraya getirilmiş çeşitli faktörlerin ahenkli bir topluluğu olarak tanımlayabiliriz. İşletmeyi istihsal faktörlerinin alelâde bir yığını olmaktan kurtaran bu ahenk sevk ve idare fonksiyonunun esasını teşkil eder ve basit bir şekilde «İşletmenin her türlü pazarlama problemine bir çözüm yolu bulmak için sarfedilen çaba» olarak tanımlayabileceğimiz pazarlama araştırmasının sevk ve idare için temel fonksiyonu «karar almada yardımcı olma»dır.

Bir ana kütleden tesadüfen çekilen n birimli bir örnekte muayyen bir vasfı haiz birimlerin toplamı $x = \Sigma x_i$ ve örnek toplamı $y = \Sigma y_i$ bulunmuş ise; $r = x/y$ şeklinde elde edilecek bir numune değeri, $R = X/Y$ parametresinin bir tahminidir. Pek tabiidir ki örnek toplamları x ve y'nin ana kütle değerleri X ve Y'nin tarafsız bir tahminleri olabilmeleri örnekleme modeline bağlıdır.

Konuyu basit bir misalle şöyle izah etmek mümkündür. Belirli bir piyasada belirli bir dönemde muayyen bir cins malın satışları muhtelif markaların toplamı $Y = A + B + C + \ldots$ olarak tahmin edilmiş olsun. Aynı piyasada aynı dönemde firmamız mamûlü A malının satışı X'de tahminen biliniyorsa $R = X/Y$'ye mamûlün piyasa hissesi adı verilecektir.

$$R = \frac{X}{Y} = \frac{\text{A malı satışı}}{\text{A+B+C+ \ldots. satışları}}$$

Şimdi bu piyasadan alınacak bir örnek yardımıyla tahmini bir değer olan R tahmin edilmek istense; sırasıyle x ve y'ler bulunacak, elde edilecek $r = x/y$ değeri R parametresinin bir tahmini olacaktır. Ancak bu tahmin işleminde bir taraflılığın varlığı bilinmekte, bunu giderme veya azaltma yolunda bazı prensip ve kaideler vazedilmektedir. Aşağıda Var(r)'nin hesabında ayrıca bu temayülün ihmali için gerekli şartlardan kısaca bahsedilecektir ama herşeyden önce şu bilinmelidir ki bu taraflılık örnek büyüklüğü arttıkça azalacaktır.

Burada problem, R'yi tahmin etmek değil, r'yi R'nin iyi bir tahmini kabul ettikten sonra bundan hareketle ana kütle içinde aradığımız vasfı haiz birimlerin toplamını tahmin etmek olacaktır. X ve Y arasında bir bağlılık olduğu biliniyorsa, ki genellikle muayyen bir piyasada aynı cins malların toplam satışları ile tek bir malın satışı arasında -müsavi olmayan mevsimlik ve arızi dalgalanmalar hariç- büyük ölçüde bir bağlılığın olduğu söylenebilir. Y'nin gerçek değerinin bazı kaynaklardan hareketle iyi bir tahmininin yapılabilmesi halinde $\hat{X} = r \, . \, Y$ hesaplanabilecektir.

Metodun tatbikattaki diğer bir uygulaması da farklı iki tarihte yapılan satış tahminleri için elde edilen örnek sonuçlarının mukayesesidir. Bu satış tahminlerinin aynı piyasada aynı birimlerin cevaplarından derlenmiş olması halinde bulunacak r değeri ilerideki bir tarihte -eğer iktisadi ortamda büyük bir değişiklik olmamışsa- ana kütle için yapılacak tahminin güvenirliliği hakkında verilecek kararda kullanılabilecektir. Benzer bir çalışma iki farklı tarihte aynı şahıslara uygulanmış aynı soru kâğıdı sonuçlarının mukayesesi için de yapılabilir; çünkü aynı birimlerin aynı soru kâğıdına farklı tarihlerde verdikleri cevaplar arasında yüksek bir nisbi münasebetin olması beklenir.

$\hat{X} = r \, . \, Y$ şeklinde bulunacak bir toplamın variyansı r'nin variyansını Y^2 ile çarpmak suretiyle elde edilecektir. O halde evvelâ Var(r)'yi bulalım.

$$y - Y = \Delta y \qquad x - X = \Delta x$$

$$\text{Var}(x) = E(\Delta x^2) \qquad \text{Var}(y) = E(\Delta y^2) \qquad \text{ve}$$

$$\text{Kov}(x, y) = E(\Delta x \, . \, \Delta y)$$

yazılabilir.

Var(r)'nin ancak

a. Y'nin çok büyük

b. y'nin sıfıra yakın çıkması ihtimalinin ihmal edilebilir derecede küçük olması şartlarıyle hesaplanabileceği, bu şartların ise genellikle bir çok araştırmalarda gerçekleştiği kabul edilmektedir.

$$r - R = \frac{x}{y} - \frac{y}{y} R = \frac{\Delta x + X - R(\Delta y + Y)}{y}$$

$$= \frac{\Delta x - R\Delta y}{y}$$

$$y = Y + \Delta y = Y + Y\frac{\Delta y}{Y} = Y\left(1 + \frac{\Delta y}{Y}\right) \quad \text{den}$$

$$r - R = \frac{\Delta x - R\Delta y}{y}$$

yerine yaklaşımı olarak

$$= \frac{\Delta x - R\Delta y}{Y}\left(1 + \frac{\Delta y}{Y}\right)^{-1}$$

yazabiliriz. $E(r - R)^2$ yi hesaplamak için kesrin yalnız payını düşünelim.

$$E(\Delta x - R\Delta y)^2 = E(\Delta x^2) + R^2 E(\Delta y^2) - 2R\, E(\Delta x \,.\, \Delta y)$$

$$= \mathrm{Var}(x) + R^2 \,.\, \mathrm{Var}(y) - 2R\, \mathrm{Kov}(x, y)$$

$$\mathrm{Var}(r) = E(r - R)^2 = \frac{1}{Y^2}\left[\, \mathrm{Var}(x) + R^2 \,.\, \mathrm{Var}(y) - 2R\, \mathrm{Kov}(x, y) \,\right]$$

elde edilir. Dikkat edilirse yukarıda

$$\frac{\Delta x - R\Delta y}{y} \quad \text{ifadesi yerine} \quad \frac{\Delta x - R\Delta y}{Y}$$

ifadesini kullanmıştık. Bu alınış tabiatiyle variyansın hesabında bir taraflılık, bir temayül yaratır. Genellikle bu temayülün x veya örnek mevcudunun büyümesiyle azalacağı kabul edilmekte, basit bir prensip olarakta $C_y = \sigma_y/Y$ kıymetinin 0.1 den küçük olması öngörülmektedir; çünkü B temayülü gösteriyorsa $\sigma^2 = \sigma^2 (1 + B^2)$ de B'nin σ^2 üzerindeki tesiri bu takdirde 0.04 den küçük kalmaktadır. Bu taraflılığın r ve y arasında bir ilişki olmaması halinde gözden kaybolacağı söylenebilir.

Gerçekten

$$\text{kov}(r, y) = E\,[\,(r - Er)\,(y - Ey)\,]$$

$$= E(ry - yEr - rEy + Er\,.\,Ey)$$

$$= Ex - Ey\,.\,Er - Er\,.\,Ey + Er\,.\,Ey$$

$$= X - Y\,.\,Er = -Y(Er - R) = -Y\,.\,\text{temayül}(r)$$

$$\text{kov}(r, y) = -Y\,.\,\text{temayül}\,(r)$$

$$\text{temayül}(r) = -\text{kov}(r, y)/Y$$

$$= -\frac{\varphi_{ry}\,\sigma_r\,\sigma_y}{Y}$$

$$= -\varphi_{ry}\,.\,C_y\,.\,\sigma_r \quad \text{veya}$$

$$\frac{|\text{temayül}\,(r)|}{\sigma_r} \geqq C_y \qquad \text{yazılabilir.}$$

O halde temayül örnek büyüklüğü arttıkça veya r ile y arasındaki bağ zayıfladıkça azalacaktır. Tatbikatta bu taraflılığın ihmali

$$(1 - f)\,\frac{s_x^2}{n\bar{x}^2} \qquad \text{veya} \qquad (1 - f)\,\frac{s_y^2}{n\bar{y}^2}$$

nin 0.05'den küçük olması şartına bağlanmakta, x ve y arasında nisbi bir bağlılığın bulunması hallerinde -ki bu özellik piyasa araştırmalarında x'in firma satışlarını, y'nin bu satışlar dahil tüm satışlarını temsil ettiği problemler için geçerlidir- bu şart 0.15 şeklinde hafifletilmektedir.

$$\text{var}(r) = \frac{1}{y^2}\,[\ \text{var}(x) + r^2\,\text{var}(y) - 2r\,\text{kov}(x,y)\ ]$$

ifadesi

$$\text{var}(x) = n^2\,\text{var}(\bar{x}) = n\,.\,s_x^2 \quad \text{den hareketle}$$

$$\text{var}(r) = \frac{n}{y^2}\,(s_x^2 + r^2\,.\,s_y^2 - 2r\,.\,\varphi_{xy}\,.\,s_x\,.\,s_y)$$

$$= \frac{1}{n\bar{y}^2}\,(s_x^2 + r^2\,s_y^2 - 2r\,.\,\varphi_{xy}\,.\,s_x\,.\,s_y) \quad \text{veya}$$

$$= \frac{1-f}{n\bar{y}^2}\,(s_x^2 + r^2\,s_y^2 - 2r\,.\,\varphi_{xy}\,.\,s_x\,.\,s_y)$$

yazılabilir. Buradaki φ_{xy}; x ve y arasındaki korelasyon katsayısıdır.

$$\varphi_{xy} = 1 \quad \text{için} \quad \sigma_r^2 = r^2\,\frac{1-f}{n}\,(\frac{s_x}{\bar{x}} - \frac{s_y}{\bar{y}})^2$$

ile variyans(r) minimum olur. Genellikle x ve y arasında oldukça yüksek bir münasebetin varlığı söz konusu edilebilir. Eğer x ve y farklı örneklerden elde edilmişlerse, bu durumda $\varphi_{xy} = 0$ olacağından örnekleme variyansı oldukça yüksek çıkar ve variyans(r) maksimum değerine $\varphi_{xy} = -1$ halinde erişir.

$$\sigma_r^2 = r^2\,.\,\frac{1-f}{n}\,(\frac{s_x}{\bar{x}} + \frac{s_y}{\bar{y}})^2$$

Buraya kadar hep var(r)'den bahsettik; halbuki bize lâzım olan Var(rY)'dir. Bu ise daha önce de belirtildiği gibi r'nin variyansının Y^2 katına eşittir.

$$\text{Var}(X) = \text{Var}(rY) = Y^2\,.\,\frac{1-f}{n\bar{y}^2}\,(s_x^2 + r^2\,s_y^2 - 2r\,.\,\varphi_{xy}\,.\,s_x\,.\,s_y)$$

$$= N^2 \bar{y}^2 \frac{1-f}{n\bar{y}^2} (s_x^2 + r^2 s_y^2 - 2r \cdot \varphi_{xy} \cdot s_x \cdot s_y)$$

$$= N^2 \cdot \frac{1-f}{n} (s_x^2 + r^2 s_y^2 - 2r \cdot \varphi_{xy} \cdot s_x \cdot s_y)$$

bu ifadeyi

$$\mathrm{Var}(\hat{X}) = N^2 \cdot \frac{1-f}{n} s_x^2 + N^2 \cdot \frac{1-f}{n} \left[r(r \cdot s_y^2 - 2 \varphi_{xy} s_x \cdot s_y) \right]$$

şeklinde yazar ve $r = 0$ koyarsak doğrudan

$$\mathrm{Var}(\hat{X}) = N^2 \cdot \frac{1-f}{n} s_x^2\text{'e}$$

yani ortalamadan hareketle elde edilen toplamın variyansına varırız. O halde

$$\mathrm{Var}_r(\hat{X}) = \mathrm{Var}_0(\hat{X}) + N^2 \cdot \frac{1-f}{n} \left[r(r \cdot s_y^2 - 2 \varphi_{xy} s_x \cdot s_y) \right]$$

yazılabilir. Buna göre oran - toplam metodu ile elde edilen $\mathrm{Var}_r(\hat{X})$ nin ortalama - toplam metodu ile elde edilen $\mathrm{Var}_0(\hat{X})$ den küçük olabilmesi

$$r(r \cdot s_y^2 - 2 \varphi_{xy} s_x \cdot s_y)$$

nin negatif olabilmesine bağlıdır. Tatbikatta genellikle φ_{xy} bire çok yakın olduğu ve aynı örnekten elde edilen s_x ile s_y birbirinden çok büyük fark etmeyeceği için

$$r \cdot s_y^2 < 2 s_x \cdot s_y$$

çıkmakta ve neticede

$$r(r \cdot s_y^2 - 2 \varphi_{xy} s_x \cdot s_y)$$

negatif değer kazandığından

$$\mathrm{Var}_r(\hat{X}) < \mathrm{Var}_0(\hat{X})$$

bulunmaktadır[2].

O halde kısaca ifade edilirse, toplamın variyansını küçültmek için oran - toplam metodu, ortalama - toplam metodu yerine tercih edilmeli, Y'nin gerçek kıymetinin hesaplanabildiği veya çok iyi bir tahminin yapılabildiği durumlarda $\hat{X} = N \cdot \bar{x}$ yerine $\hat{X} = r \cdot Y$ ifadesi kullanılmalıdır.

V.

SONUÇ

Modern istatistik metodları, iktisadi modellerdeki parametrelerin ölçülmesinde büyük bir yardımcıdır. Özellikle sosyal olayların ve işletmecilik alanındaki problemlerin incelenmesinde önem kazanan karar verme tekniğinde daima bir risk «belirsizlik» sorunu ile karşılaşılır. İlim adamı veya araştırıcı az veya çok bu risk ile birlikte yaşamasını öğrenir. Hesaplanabilen risklerin hesaplanamayan risklere tercihi, ayrıca hesaplanabilmesi yanında istenildiği kadar küçük tutulabilmesi imkânlarının da araştırılması neticede ihtimal teorisini ve ihtimali örnekleme metodolojisini geliştirmiştir.

Örnek üzerinde tahminin, tahmin edilmek istenen gerçek değere tam sayımdan daha yakın olması hususu küçük bir ihtimal olmaktan

---

2) x ve y'nin zümrelere göre sondaj metodu ile elde edilmiş örnek değerleri olmaları halinde her zümre için r_h hesaplanmalı ve

$$r = \sum \frac{Y_h}{Y} \cdot r_h$$

örnek değeri olarak alınmalıdır. Bu takdirde

$$\mathrm{Var}(r) = \frac{1}{Y^2} \sum N_h^2 \frac{1 - f_h}{n_h} (s_{hx}^2 + r^2 s_{hy}^2 - 2r\, \varphi_{xy}\, s_{hx} \cdot s_{hy})$$

olacaktır. Bu tahminde de taraflılık r_h ların temayülüne bağlıdır; o da n_h ların artmasıyla azalacaktır.

çok uzak bulunduğundan bu teknik oldukça geniş ölçüde kullanılmakta, örnek değerlerinden hareketle ana kütle parametrelerine varış ve bu tahminin geçerlilik sınırları çeşitli formüllerin uygulanma sahasını teşkil etmektedir.

İşletme idaresinde ortaya çıkan hadiseler ekonomik bir mahiyet taşırlar. Ekonomik hadiselerin değişken karekterinin zaman ve mekân içinde tam bir tarifinin yapılmasına, hudutlarının kesin bir şekilde çizilmesine imkân yoktur. Bu gibi değişkenler daima tesadüfi bir takım tesirlerin altında kalırlar. İşletmelerde faaliyet neticeleri muhasebe kayıtlarında yer alır ve genellikle bu eski kayıtlardan çıkarılan analitik ifadeler işletmenin ilgili hadisesine ait modelini teşkil eder. Ancak çoğu defa tetkik edilen hadisenin karekterinin değişkenine ait bilgilerin geçmiş faaliyet dönemlerinde elde edilememiş olduğuna rastlanır. Böyle bir durumda analitik model tesisi mümkün olamıyacağından geleceğe ait tahmine girişilemiyecektir. Bu gibi hallerde bilgileri aynı faaliyet dönemi içinde yapılacak müşahedeler yardımıyla temin etmek gerekmekte, müşahede ise belli bir örnek üzerinden yapılacağından örnekleme modelinin tesbiti, örneğin seçimi, neticelerin takdiri ve hata payları gibi hususlar istatistik metodların kullanılmasına dayanmaktadır.

İşte bu kısa etüd bir ihtimali sondaj örneğinden hareketle ana kütle toplamının tahmininde çok sık kullanılan bir metod yerine, işletmeler için bazı şartlarla geçerli ve daha fazla uygulanma imkânına sahip başka bir metodun hatırlatılması amacıyle hazırlanmıştır.

M. Slonim (Terc. F. Akün) «Numune Alma Tekniği Esasları» D.İ.E. 1966.

H. Arkın - R.R. Colton (Tere. S. Kendir) «İstatistik Metodları» Ankara 1968.

E.J. Cogan - Ö.G. Kemeny (Terc. N. Uzgören) «Modern Matematik Metodları ve Modelleri» İstanbul 1965.

B.W. Gnedenko - A.J. Chintschin (Terc. L. Birand) «İhtimaller Hesabına Giriş» İstanbul 1963.

R. Goodman «Statistics» The English Universities Press London 1967.

L. Kish «Survey Sampling» John Wilcy London 1967.

A.J. Merrett - G. Bannock «Business Economics and Statistics» Hutchinson Ltd. London 1962.