

T.C.
MİMAR SİNAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI
YÜKSEK LİSANS PROGRAMI

78389

ZÜMRELERE GÖRE ÖRNEKLEMEDE REGRESYON YÖNTEMİ

YÜKSEK LİSANS TEZİ

Hazırlayan: Meral YAY / 94111038

Danışman Öğretim Üyesi: Prof. Dr. M. Kemal Yoğurtçugil

İstanbul, 1998

T.C. YÜKSEKÖĞRETİM KURULU
BÖLGE BAŞKANLIĞI MERKEZİ

İÇİNDEKİLER

1. GİRİŞ	1
2. İSTATİSTİKSEL BİR TEKNİK OLARAK ‘ÖRNEKLEME’	3
3. TAHMİN YÖNTEMLERİ	15
3.1. Momentler Yöntemi	15
3.2. Maksimum Likelihood Yöntemi	16
3.3. Regresyon Yöntemi	17
3.4. Nisbet Yöntemi	23
4. ÖRNEKLEME YÖNTEMLERİ	25
4.1. Basit Tesadüfi Örneklem	25
4.1.1. Örnek Ortalamaları Teorik Bölünmesi	27
4.1.2. Örnek Varyansları Teorik Bölünmesi	30
4.1.3. Örnek Nisbetleri Teorik Bölünmesi	31
4.2. Zümrelere Göre Örneklem	34
4.3. Kümelere Göre Örneklem	34
4.4. Sistemik Örneklem	38
5. ZÜMRELERE GÖRE ÖRNEKLEME	41
5.1. Momentler Yöntemi İle Tahmin	45
5.1.1. Anakütle Ortalamasının Tahmini	45
5.1.2. Anakütle Nisbetinin Tahmini	50
5.2. Regresyon Yöntemi İle Tahmin	52
5.2.1. Anakütle Ortalamasının Ayrı Tahmini	54
5.2.2. Anakütle Ortalaması Ayrı Tahmininin Varyansı	55
5.2.3. Anakütle Ortalamasının Bileşik Tahmini	56
5.2.4. Anakütle Ortalaması Bileşik Tahmininin Varyansı	57
5.2.5. Zümrelere Göre Örneklemde Regresyon İle Tahmin Yöntemlerinin Karşılaştırılması	57
6. UYGULAMA	61
7. SONUÇ	79
KAYNAKLAR	81

1. GİRİŞ

Günümüzde istatistik birçok alanda kullanılmaktadır. Sayısal verilerin girdiği her türlü bilimsel çalışmada yeralan istatistik, bilgisayar teknolojisinin ilerlediği şu günlerde yerini gitgide sağlamlaştırmaktadır.

İstatistik, sayısal olarak ifade edilebilen verilerin toplanması, düzenlenmesi ve çeşitli yöntemlerle analiz edilerek elde edilen sonuçların yorumlanmasını içeren bir metod veya metodlar topluluğudur.

İstatistiğin amacı bilinen verilerden yararlanarak bilinmeyenler hakkında belli bir hata oranında yorum yapmak ve kullanılmak üzere sunmaktır.

İstatistiğin temel konularından biri de ‘Örnekleme’dir. Üzerinde araştırma yapılan anakütle çok geniş bir alana yayılmış olabilir. Bu durumda oluşturulacak örnek veya örneklerden yararlanarak, belli bir hata payı ile anakütle hakkında bilgi edinilebilir. Bu şekilde istenilen bilgiye daha çabuk ve daha az masraf ile ulaşılabilir.

Yapılan bu tez çalışmasında, öncelikle örneklemenin ne denli gerekli bir teknik olduğu üzerinde durulmuş ve örneklemeyle ilişkin temel bilgiler verilmiştir. Daha sonra anakütle parametrelerine ilişkin nokta tahmin yöntemleri anlatılmıştır. Bir sonraki bölümde temel örnekleme yöntemlerinden bahsedilerek, son bölümde; tahmin yöntemlerinden ‘Regresyon Yöntemi’ ile örnekleme yöntemlerinden ‘Zümrelere Göre örnekleme Yöntemi’ birlikte ele alınmıştır.

Çalışmanın amacı, en sık kullanılan örnekleme yöntemi olan Basit Tesadüfi Örnekleme yerine Zümrelere Göre Örnekleme yöntemini kullanmanın faydasını görmek ve bu yöntemle birlikte ele alınan nokta tahmin yöntemi olan ‘Regresyon Yöntemi’ nin diğer tahmin yöntemlerine sağladığı üstünlüğü belirtmektir.

Tez çalışmam süresince; tecrübe, bilgi ve hoşgörüsünü benden esirgemeyen değerli hocam Prof.Dr.M.Kemal Yoğurtçugil'e teşekkürü bir borç bilirim.

İstanbul,1998

Meral YAY

2. İSTATİSTİKSEL BİR TEKNİK OLARAK ‘ÖRNEKLEME’

Bilgimiz, davranışlarımız ve faaliyetlerimiz genelde çok büyük alanlara yayılmış olmakla beraber elde edebileceğimiz verilerin yapısı ve uygulamada karşımıza çıkabilecek sorunlara ilişkin çözümler bazı kısıtlara dayandırılır. Her zaman çok geniş çaplı araştırmalar yapmak ve bu araştırmalara dayanarak amaca göre bir yoruma varmak dışarıdan görüldüğü kadar kolay değildir. Bu noktada istatistiksel tekniklere başvurmak bir gereksinim haline gelmiştir. Tabi ki istatistiksel teknikler kullanılırken , matematik bilgisi ile de desteklenmektedir. İstatistiğin matematik ile olan ilişkisi diğer bilim dalları ile olan ilişkisinden çok daha yakındır. İstatistik, matematikten aldığı yöntemleri işleyerek toplu olayların incelenmesine yarayacak şekle getirmekte ve kullanıma sunmaktadır.

Araştırmalarda gerekli olan bilgi ve veriler, ya konu ile ilgili birimlerin tümünden ya da tamamı temsil ettiği varsayılan bir grup birimden sağlanır. Araştırması yapılacak konu ile ilgili verilerin toplanması sırasında kullanılan teknikler temel olarak iki grupta toplanır. Bunlar tamsayım ve örneklemedir. Bazı veriler sadece tamsayım veya sadece örnekleme tekniği ile elde edilebilir. Anakütlenin yani araştırmacının ilgilendiği ortak özellikleri olan birimlerden oluşan yığının çok önem arzetmeyen bir kısmı için bilgi gerekiyorsa, çalışmanın tamsayım ile yapılması şarttır. Hakkında bilgi edinilmek istenen topluluğa dahil birimlerin tümünün incelenmesine tamsayım denilirken, bir anakütleden amaca uygun olarak rasgele seçilmiş ve daha az sayıda birimden oluşan verilere dayanarak anakütle hakkında genel yargılara

varma işlemine ise örnekleme denilir¹. Tamsayım veya örneklemeden hangisinin tercih edileceği, araştırmacının hedef aldığı doğruluk derecesine ve elindeki olanaklara bağlıdır.

İnceleme konusu olacak araştırma birimleri ülkeler, firmalar, yöneticiler, işçiler, aileler, tüketiciler ve benzeri birimler olabilir. Bu birimlerin niteliğinden de anlaşılacağı gibi genellikle anakütller çok sayıda birimden oluşmaktadır. Bu neden başta olmak üzere firmaların araştırmaya ayıracak zaman ve finansal kaynaklarının kısıtlı olması nedeniyle çoğu zaman örnekleme yöntemi tamsayım yöntemine tercih edilmektedir. Bunun yanında tamsayım ile örneklemenin birlikte kullanıldığı durumlar da mevcuttur. Bunlar aşağıdaki gibi sıralanabilir.

- Tamsayım sonucu elde edilen bilgileri kısa sürede derli toplu sunmak amacına yönelik fakat uygulaması zorunlu olmayan örnekleme,
- Tamsayımın doğruluğunu kontrol ve elde edilen bilgilerin kalitesini değerlendirmek için yapılan örnekleme,
- İki sayım arasında eskiyen verileri yenileme amacına yönelik yapılan örnekleme².

İster tamsayım ister örnekleme yöntemi ile elde edilsin istatistikler çok sayıda hatayı da beraberinde getirir. Bu hataları aşağıdaki gibi sıralamak mümkündür:

- Amacın ve gereksinim duyulan bilgilerin belirlenmesindeki başarısızlık.
- Araştırma araçlarının seçimi veya düzenlenmesindeki başarısızlık.

¹ Yoğurtçugil,M.K., Örnekleme, İstanbul: Sernet Matbaası, 1976, s.5.

² Serper,Ö., Aytaç, M., Örnekleme, İstanbul: Filiz Kitabevi, 1988, s.4.

•Toplanacak bilgilerin niteliğine uygun gelen kimselerin seçiminde yapılan başarısızlık.

- Anakütlenin tanımındaki başarısızlık.
- Tanım ve açıklamalardaki başarısızlık.
- Yanıtsız bırakmalardan doğan hatalar.
- Yanıtların geç gelmesinden doğan hatalar.
- Yanlış verilen yanıtlar.
- Farkında olmadan değişik şekilde yanıtlamadan doğan hatalar.
- Sayım memurlarının neden olduğu yanıt hataları.
- Sayım veya örneklemeyi uygulayan kimse veya kurumun neden olduğu hatalar.
- İşleri düzenleme ve yönetmede yapılan başarısızlık.
- Sayım ve örnekleme tarihinin veya bilgilerin ilişkin bulunduğu devrenin yığını iyi temsil etmemesinden doğan hatalar.
- Sınıflama planlarının iyi hazırlanmamasından doğan hatalar.
- İşlem hataları.
- Sonuçların yanlış yayımlanması.
- Örnekleme hataları.

Örnekleme hatalarını biraz daha geniş bir biçimde ele almak yerinde olur. Örnekten hesaplanan değer ile anakütle parametresinin birbirine eşit çıkması genellikle karşılaşılmayan bir durumdur. Örnek değeri ile anakütle parametresi arasındaki farka toplam hata denir. Örneklemede önemli sayılabilecek iki tip hata vardır. Bunlar sistematik hata ve tesadüfi hatadır. Toplam hata ile sistematik ve tesadüfi hata arasındaki ilişki aşağıdaki gibi ifade edilebilir:

$$\text{ToplamHata} = \sqrt{(\text{SistematikHata})^2 + (\text{TesadüfiHata})^2}$$

Sistematik hata, örnek değeri anakütle parametresinden tek yönlü sapma eğilimi gösterdiğinde ortaya çıkar. Sistematik bir hata yapıldığında, örnek değeri anakütle parametresinden ya hep küçük ya da hep büyük çıkar. Anakütleden çok sayıda n birimli örnek çekildiği ve bunların mesela aritmetik ortalamaları hesaplandığı zaman ortalamaların birbirinden ve anakütle ortalamasından farklı olduğu görülür. Bu farka da tesadüfi hata denir. Tesadüfi hatalar her iki yönde etkili oldukları için birbirlerini yok ederler.

Anakütle ile anakütleden çekilen örnekler arasındaki ilişkiyi inceleyen ve çok geniş bir uygulama alanına sahip olan, aynı zamanda iki farklı örneğe ait olan örnek değerleri arasında gözlenen farkların tesadüfen mi yoksa gerçekten anlamlı mı olduğunun saptanmasında da kullanılan örneklemenin, neden gerekli olduğundan bahsetmek bundan sonra yapılacak açıklamalara da ışık tutacaktır. Örneklem seçilerek yapılan incelemelerin tümü maliyet ve zaman açısından ekonomik olduğu kadar aynı zamanda tamsayım yapıldığında elde edilen sonuçlar kadar güvenilir ve sağlıklıdır. İncelenecek anakütle genişledikçe araştırmacının her birimi tek tek incelemesi de güçleşecektir. Yani yığının büyümesi maliyet artışı ile doğru orantılı bir hale gelecektir. Bu yüzden her birimin tek tek incelenmesi yerine, örneklem kullanmak suretiyle hem maliyet azaltılır hem de araştırmacı için büyük bir zaman kazancı elde edilir. Ayrıca araştırma süresince elde edilmesi amaçlanan pratik sonucun mümkün olan en kısa zamanda elde edilmesi de örnekleme yöntemlerinin kullanılması ile sağlanır. Yıllar boyu süren araştırmalar sonuçlandığında büyük bir olasılıkla araştırmaya neden olan sorunlar ortadan kalkmış olabilir. Çevresel etkilerinde sürekli değişen bir yapıya sahip olması araştırmanın örnekleme tekniğine dayandırılmasını zorunlu kılar. Araştırmada kullanılacak eleman sayısının yüksek olması da araştırmanın gerçekleşmesinde büyük rol

oyunar. Bu ise hem maliyeti etkiler hem de konuda eğitilmiş insan gücü sorununu gündeme getirir.

Örneklemeyi zorunlu kılan bir başka neden de bazı araştırmaların denekler üzerinde zararlı etkilere yol açma olasılığının olmasıdır. Örneğin; herhangi bir hayvan ya da bitki türünün tamsayımla araştırılması o türün ortadan kalkmasına neden olabilir.

İnceleme konusu olan yığının tam olarak saptanmasında çeşitli güçlüklerle karşılaşılabilir. Örneğin; boşanma üzerinde çalışan bir toplumbilimci ancak kayıtlara geçmiş olan boşanmaları içeren bir yığın oluşturabilir. Halbuki, ayrı yaşama, terk edilme gibi durumlarda da gerçekten boşanma sonucu karşılaşılan koşullar geçerli olmaktadır. Ancak ayrı yaşama ve terk edilme durumlarının kayıtlarda yeralması sözkonusu olmadığı için bu birimlerin yığına dahil edilmeleri olanaksızdır.

Örnekleme çalışmasına başlamadan önce araştırma iyi bir şekilde planlanmalı ve örnekleme bu plana göre yürütülmelidir. Bir örnekleme araştırması yapılırken bazı temel adımların dikkate alınması araştırma açısından büyük önem taşır. Örnekleme işlemine başlarken mutlaka bir amaç gözönünde bulundurulur; herhangi bir ürünün piyasası hakkında ayrıntılı bilgi edinmek veya piyasaya sürülecek olan yeni bir ürünün tutup tutmayacağı hakkında araştırma yapmak sözkonusu olabilir. Örneklemenin amacı da bu araştırmayı fazla maliyet gerektirmeden en kısa sürede sonuçlandırmaktır. Herşeyden önce araştırmanın amacının kolay anlaşılır bir şekilde ifade edilmesi ve hedef kütlenin belirlenmesi gerekir. Bazı durumlarda içinden örnek alınan anakütle, yani 'örnekleme anakütlesi' ile hakkında bilgi toplanmak istenen anakütle yani 'hedef anakütle' birbirinden farklıdır. Bu iki anakütlenin aynı olmasına dikkat edilmelidir. Eğer araştırmayı kolaylaştırmak amacıyla örnekleme anakütlesi, hedef anakütleden daha dar tutulmuş ise sonuçların sadece örnekleme anakütlesine ait olduğu unutulmaması gereken

bir husustur. Böyle durumlarda iki anakütle arasındaki farkların özellikleri ve dereceleri hakkında ek bilgi edinilmesi yerinde olur. Bu aşama olmaksızın yani hedef anakütle ve örnekleme anakütlesi belirlenmeksizin araştırmanın planlanması zorlaşır ve karar aşamasında karışık bir durumla karşılaşmak kaçınılmaz olur. Ele alınacak anakütlenin saptanması ve tanımlanması örnekleme planının temel adımlarından biridir. Anakütle tanımlandıktan sonra yapılması gereken araştırma için veri toplanmasıdır. Toplanacak verilerin araştırmanın amacına uygun olmasına ve gereksiz verilerin gözardı edilmesine dikkat edilmelidir. Anakütle hakkında belkide hiç kullanılmayacak gereğinden fazla bilgiyi toplamak, sık karşılaşılan bir sorun haline gelmiştir. Bunun bir sonucu olarak fazla bilgi elde etmeye çalışırken hiçbir güvenilirliği olmayan bilgilerle yüzyüze kalmak kaçınılmaz olur. Toplanacak verilerin niteliği, önemli sorulara olduğu kadar önemsiz sorulara da cevap verebilecek şekilde olmalıdır. Ancak bunu yaparken örnekleme amacının gözden kaçmamasına ve fazla ayrıntıya girmemeye özen gösterilmelidir. Bu aşamalardan sonra söz edilmesi gereken bir husus da; araştırmanın güvenilirliğini sağlamaktır. Örnekleme sonuçları bir takım istatistiksel testlere de tabi tutulacak ise, güvenilirlik derecesinin önceden belirlenmesi, sonuçların istenilen şekilde yorumlanabilmesine imkan sağlayacaktır. Anakütle tanımlanması ve veri toplama aşamasından sonra araştırma sırasında kullanılacak örnekleme yöntemi belirlenmelidir. Araştırmanın amacına, maliyet,süre v.b. kısıtlara ve anakütlenin özelliklerine göre örnekleme yöntemlerinden en uygun olanının seçilmesi gerekir. Bazı durumlarda birkaç örnekleme yöntemi birarada veya birbirini takip eden bir şekilde kullanılabilir. Daha önce de ifade edildiği gibi ister tamsayım ister örnekleme yöntemi kullanılsın, araştırma mutlaka bir hata payı taşır. Bu hatayı tamamen ortadan kaldırmak mümkün değildir. Bu yüzden belli bir hata payının örnekleme planı çerçevesinde belirlenmesi gerekir. Genellikle bu hata payı 0.01 veya 0.05 olarak seçilir. Bu, tahmin edilen anakütle parametrelerinin %99 veya %95 olasılıkla oluşturulan güven

aralıkları içinde bulunacağı anlamına gelir. Hata payı da belirlendikten sonra çerçeve belirlenerek araştırmada bir adım daha ileri gidilir. Bu arada çerçeve ile ilgili tanımı da kısaca vermekte fayda vardır; çerçeve, birime ait her türlü bilginin yani birimin adı, soyadı, cinsiyeti, mesleği, kan grubu vb. gibi bilgilerin oluşturduğu bir listedir. Çerçevenin düzenlenmesi çok maliyetli olabilir. Bu yüzden ne tür bir çerçevenin en ekonomik olacağı önceden belirlenmelidir. Çerçevenin mümkün olduğunca anakütlenin büyük bir kısmını kapsayacak şekilde olması tercih edilir, ancak maliyeti de gözönünde bulundurarak anakütlenin önemsiz kısımları kapsam dışı bırakılır. Bunlara ilaveten örneklem büyüklüğü ile anakütle büyüklüğü arasındaki oran olarak tanımlanan örneklem oranının bulunması da örneklem planında yer alır. Başka bir deyişle, anakütle birimlerinin yüzde kaçının örneklem girdiğinin belirlenmesinde yardımcı olan örneklem oranının bilinmesi gerekir. Anakütle parametresini en doğru şekilde tahmin edebilecek örneklem yöntemlerinden biri seçilerek örneklem planını oluşturmaya devam edilir. Bu tekniklerin en önemlisi Basit Tesadüfi Örneklemedir. Bu yöntemin en önemli özelliği anakütle içindeki her birime çekilecek örneğe girme açısından eşit şans vermesidir. Bu şans yığına dahil her birimin hüviyetinin belirlenmesi halinde sağlanabilmekte bu ise uygulamada genelde gerçekleştirilememektedir. Bu nedenle örneklem yöntemi eşit şans yerine önceden belli şans prensibine dönmekte bu ise Sistemik Örneklem, Kümelere Göre Örneklem gibi yöntemleri gündeme getirmektedir. Örneklemde anakütle dağılması çok büyük ise örneklem planını hazırlamak ve uygulamak çok masraflı ve zaman alıcı olacağından yığının varyansını küçükmeye dayalı Zümrelere Göre Örneklem yöntemi daha çok kullanılmaktadır. Bir sonraki adımda ise örnek seçimine gidilir. Örnek seçiminde temel kural; örneğin olasılık kurallarına dayanarak çekilmesi, böylece anakütledeki her birime örneğe girme açısından eşit şans tanınmasıdır. Sıra örneklem planının en önemli aşamasında yani tahmindedir. Anakütle parametresi, anakütlenin içinden çekilen örnek

değerinden hareketle iki şekilde tahmin edilebilir. Bunlar ‘Nokta Tahmin’ ve ‘Aralık Tahmin’dir. Anakütle parametresi ile örnek değeri arasındaki fark ne kadar küçük ise tahminin o derece başarılı olduğu söylenebilir. Nokta tahmin, örnek değerinden hareketle anakütle parametresini tek bir değer olarak tahmin etmektir. Bu tahmin tabiki yaklaşık bir tahmindir. Nokta tahminin iyi bir tahmin olduğunu gösteren bazı özellikleri de bu arada açıklamak da fayda vardır. İyi bir nokta tahmin yansız(tarafsız), tutarlı(uygun), etkin ve yeterli olmalıdır.

İyi bir nokta tahminin yansız olması demek; N büyüklüğündeki bir anakütleden n birimli C_n^N tane örnek çekildiği ve bu örnek değerlerinin herbiri için bir örnek değeri hesaplandığı düşünülürse, örnek değerleri teorik bölünmesinin beklenen değerinin anakütle parametresine eşit olması demektir. Yani bu tahmin sistematik bir hata içermez. Mesela, örnek ortalaması $\bar{x}$ ve nisbeti p, anakütle ortalaması $\bar{X}$ ve nisbeti P'nin yansız birer tahminleridirler.

Anakütle içinden çekilen örneğin büyüklüğü arttığı zaman tahmin değerinin tahmin edilen parametrenin gerçek değerine yaklaşmasına ve örnek büyüklüğü anakütle büyüklüğüne eşit olduğu durumda ise tahmin değeri ile gerçek değer in çakışmasına nokta tahminin tutarlılık özelliği denir. Tutarlılık özelliği; λ çok küçük bir pozitif sayı olmak koşulu ile aşağıdaki gibi ifade edilir:

$$\lim_{n \rightarrow \infty} P\{|\bar{x} - \bar{X}| < \lambda\} \rightarrow 1 \quad (2.1)$$

Anakütle parametresinin tahmininde sistematik hata taşımayan birden fazla örnek değerinden hangisinin örnek değerleri teorik bölünmesi varyansı küçük ise , o nokta tahminin etkin bir tahmin olduğu söylenir. Mesela, örnek ortalamaları teorik bölünmesinin varyansı örnek medyanları teorik bölünmesinin varyansından daha küçüktür. Bu yüzden anakütle ortalamasının tahmininde; örnek ortalaması medyana nazaran daha etkin bir tahmindir.

Yeterli bir nokta tahmin ise, hesabında tüm örnek değerlerinin kullanıldığı bir tahmindir. Örneğin, anakütle ortalamasının tahmininde örnek ortalaması hesaplanırken tüm birimler kullanıldığı için medyana nazaran daha yeterli bir tahmin olduğu söylenebilir.

Aralık tahmini ise şu şekilde açıklamak mümkündür: Anakütle parametresi, içinden çekilen örnek değeri yardımıyla bir nokta şeklinde veya bu noktayı içeren bir aralık şeklinde tahmin edilebilir. Anakütle parametresini, önceden belirlenen bir olasılık düzeyinde aralık şeklinde tahmin etmeye aralık tahmin denir. Hatırlanacağı gibi nokta tahminde anakütle ortalaması örnek değeri $\bar{x}$ ile tahmin ediliyordu. Ancak tesadüfi hatalar nedeniyle, gerçek değer ile tahmini değer arasında genellikle bir fark bulunur. Anakütle ortalaması, 'd' denilen bu farkı gözönüne alarak

$$\bar{x} - d < \bar{X} < \bar{x} + d \quad (2.2)$$

şeklinde tahmin edilirse, aralık tahmin yapılmış olur. Örnek değeri ile anakütle parametresi arasında oluşacak fark yani toplam hata, belirli bir olasılıkla belli hudutlar arasında tutulabildiği sürece tahminin geçerlilik payını arttığı söylenebilir. Bu olasılık α ile gösterilirse;

$$P(\bar{x} - d < \bar{X} < \bar{x} + d) = 1 - \alpha \quad (2.3)$$

yazılabilir. Burada ifade edilen, anakütle parametresi olan $\bar{X}$ nın geçerlilik hudutları olan $\bar{x} - d$ ve $\bar{x} + d$ yi ancak α olasılığı ile aşabileceğidir. Ne kadar dar bir güven aralığı sağlanırsa, tahmin o derece başarılı olur.

Daha önce de ifade edildiği gibi, örneklemenin amacı daha az işgücü ve maliyetle anakütlenin bilinmeyen parametreleri hakkında en doğru tahmini yapmak idi. En doğru ve güvenilir şekilde tahmin yapabilmek için ise örnekleme planının bütün aşamalarının son derece dikkatle ele alınmış olması gereklidir. Ayrıca; anakütle parametrelerinin tahmininde aranan yansızlık, yeterlilik, etkinlik ve uygunluk gibi özelliklerin de gözönünde bulundurulması

gerekmektedir. Çünkü bu özelliklerin hepsine birden sahip olan tahminlerin, bilinmeyen anakütle parametrelerine çok yakın olduğu daha önce yapılan istatistiksel çalışmalarda kesin olarak ortaya konulmuştur. Fakat bazen bu özelliklerin bir veya birkaçından fedakarlık etme gereği ortaya çıkabilir. Bu gibi durumlarda seçilen örnekleme yöntemi çok büyük önem taşır. Hangi özellikten vazgeçileceğine ise tecrübeler ve örnekleme yöntemine dayanarak karar verilir. Araştırmanın daha önceki planlara uyup uymadığı sık sık kontrol edilir, cevapların doğruluğu incelenir ve gereken yerlerde kontrol örnekleme yapılır. Örnekleme planının son aşaması ise bilgilerin düzenlenip istatistiksel analizler yapılması ve değerlendirme yaparak başka bir tarihte aynı anakütle ile ilgili yapılacak araştırmalarda kullanılmak üzere saklanmasıdır.

Yapılan örneklemenin başarılı olup olmadığının ölçüsü olarak aşağıdaki hususlar sayılabilir:

- Örnek çekilecek anakütle hakkında bazı özel bilgiler gereklidir.
- Seçme işlemi ilgilendiğimiz özellik veya değişkenden bağımsız olmalıdır.
- Örnek bir ön yargıya yervermeden ve sistematik bir farklılık yaratmayacak şekilde seçilmelidir.
- Örneğe alınan birimler birbirinden bağımsız olmalıdır.
- Verilerin seçildiği alanlarla diğer alanlar arasında temel farklılıklar bulunmamalıdır.
- Örneğe alınacak verilerin hepsine aynı şartlar istisnasız uygulanmalıdır³.

Bu koşullar altında yapılan örneklemenin başarılı bir örnekleme olduğu söylenebilir.

³ Yoğurtçugil, M.K., a.g.e., s.18.

Örnekleme yöntemi, anakütle ile anakütleden çekilmiş olan örnekler arasında mevcut olan ilişkiyi inceleyen ve çok geniş bir uygulama alanı olan bir tahmin yöntemidir. Örnekleme, anakütle ile örnek arasındaki ilişkinin yanısıra iki veya ikiden fazla farklı örneğe ait karakteristik değer arasında meydana gelen farkın tesadüfi mi yoksa gerçekten anlamlı mı olduğunu da tesbit edebilen bir yöntemdir. Bu tip soruların cevabında istatistik kararlar bakımından önemli bir yeri olan Hipotez ve Anlamlılık Testleri kullanılmalıdır.

Üzerinde araştırma ve inceleme yapılması istenen anakütleden belirlenmiş sayıda birim içeren örnek veya örneklerin meydana getirilmesi için çeşitli örnekleme yöntemleri geliştirilmiştir. Bunlar aşağıdaki gibi sıralanabilir:

- Monografi Yöntemi
- Kararlı (Temsili) Örnekleme Yöntemi
- Kota Örnekleme Yöntemi
- Basit Tesadüfi Örnekleme Yöntemi
- SistematiK Örnekleme Yöntemi
- Kümelere Göre Örnekleme Yöntemi
- Zümrelere Göre Örnekleme Yöntemi
- Latin Kare Örnekleme Yöntemi

Çalışma içinde yukarıda sayılan bu yöntemler içinden en sık kullanılanlarına yer verilecektir. Bu yüzden Monografi Yöntemi, Kararlı Örnekleme Yöntemi ve Latin Kare Yöntemlerini birer paragraf halinde açıklamak yerinde olur.

Monografi Yöntemi; anakütleye ilişkin bilginin elde edilmesi amacıyla tek bir birimin ele alınıp incelenmesidir. Ancak bu birimin, anakütlenin tamamını temsil edebilecek nitelikte olması gerekmektedir ki,

ayrıntılı bilgi elde etme olanağı verebilsin. Böylece ele alınan bu tek birim anakütlerde yeralan birimlerin ortak özelliklerine sahip olan bir birim olur.

Kararlı Örneklem, anakütle içinden çekilen birimlerin örneklem yapan kişilerin şahsi yargılarına ve görüşlerine dayanarak elde edildiği bir örneklem yöntemidir.

Latin Kare Örneklem Yöntemi ise 1926 yılında Fischer tarafından kullanılan bir yöntemdir. Uygulama alanı, genellikle üretime giren hammadde veya mamul maddenin hammaddesinin birden fazla olduğu durumlardır. Bu verilere göre düzenlenecek bir Latin Kare Tablosundan faydalanılarak işlemler yapılır.

Daha önce de belirtildiği gibi örneklemede amaç, anakütle parametrelerini örnek değerlerinden yararlanarak tahmin etmek idi. Tahmin iki şekilde yapılıyordu: aralık tahmin ve nokta tahmin. Çalışmamızda nokta tahmin yöntemlerine değinilecektir. Nokta tahmin yöntemlerinin sıralanışı aşağıdaki gibidir:

- Momentler yöntemi
- Maksimum Likelihood yöntemi
- Nisbet yöntemi
- Regresyon yöntemi

Tahmin yöntemlerine ilişkin açıklamalara çalışma içinde yer verilecek, ancak regresyon yöntemi daha ağırlıklı bir şekilde incelenecektir.

3. TAHMİN YÖNTEMLERİ

Bu bölümde örnek değerinden hareket ederek anakütle parametresini tahmin etmede kullanılan nokta tahmin yöntemlerinden bahsedilecektir.

3.1. Momentler Yöntemi

Momentler metodunda esas, x tesadüfi bir değişken ve $f(x, \theta)$ bu değişkenin bir fonksiyonu ise, fonksiyonun parametresi olan θ 'yı örneklerden elde edilen momentlere göre tahmin etmektir. Eğer x in fonksiyonunda birden fazla parametre varsa (mesela k tane), bu parametreleri tahmin etmek için k tane eşitliğe ihtiyaç duyulur. Çekilen örnekten elde edilen sıfır orijinine göre k . moment, anakütlerde aynı momentin yansız bir tahminidir. Mometler yöntemine doğrudan tahmin yöntemi de denilir.

Momentler yöntemine göre örnek ortalaması $\bar{x}$, anakütle ortalaması μ nın yansız bir tahminidir ve aşağıdaki şekilde ifade edilir:

$$E(\bar{x}) = \mu, \quad \bar{x}_i = \hat{\mu} \quad (3.1)$$

Ancak aynı eşitlik varyans için yazılamaz, yani örnek varyansı σ^2 , anakütle varyansı σ^2 nin yansız bir tahmini değildir ve

$$E(\sigma^2) \neq \sigma^2 \quad (3.2)$$

şeklinde ifade edilir. Örnek varyansları teorik bölünmesinin beklenen değeri

$$E(\sigma^2) = \sigma^2 \frac{n-1}{n} \frac{N}{N-1} \quad (3.3)$$

ile hesaplanır. N in büyüyen değerleri için N ve N-1 arasındaki farkın önemi ortadan kalkar ve ifade

$$E(\sigma^2) = \sigma^2 \frac{n-1}{n} \quad (3.4)$$

şeklini alır ki bu da $E(\sigma^2) > \sigma^2$ olduğunu gösterir.

3.2. Maksimum Likelihood Yöntemi

Amacı, bilinmeyen anakütle parametresinin en iyi tahminini veren örnek değerini bulmak olan Maksimum Likelihood Yöntemi, pratik olması ve bu yaklaşımla bulunan değer arzu edilen özelliklerin çoğuna sahip olması nedeniyle sık kullanılmaktadır. Ancak uygulamada karşılaşılan zorluk, tesadüfi değişkenlerin dağılımının bilinmemesidir.

x bir tesadüfi değişken ve bu tesadüfi değişkenin örnek değerleri $x_1, x_2, \dots, x_n$ ise likelihood fonksiyonu bu n tane değişkenin bileşik fonksiyonu olarak yazılabilir. $L(\theta) = \prod_{i=1}^n f(x_i; \theta)$ fonksiyonunu maksimum yapan θ değerinin bulmak maksimum likelihood yönteminin esasıdır. Bu fonksiyonun maksimum olabilmesi için, θ ya göre türevin sıfıra eşitlenip çözülmesi gerekir. $L(\theta)$ fonksiyonu eğer k tane parametre içeriyorsa;

$$L(\theta) = f(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n; \theta_1, \theta_2, \dots, \theta_k) \quad (3.5)$$

fonksiyonunu maksimum kılan

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_1} = 0$$

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_2} = 0$$

$$\begin{matrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{matrix}$$

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_k} = 0 \quad (3.6)$$

k bilinmeyenli k tane eşitlik çözülür.

$$x_1 \text{ in olasılık yoğunluğu} \quad P_1 = \frac{1}{\sqrt{2n}} e^{-1/2(\bar{x}_1 - \theta)^2} \quad (3.7)$$

$$\bar{x}_2 \text{ nin olasılık yoğunluğu} \quad P_2 = \frac{1}{\sqrt{2n}} e^{-1/2(\bar{x}_2 - \theta)^2} \quad (3.8)$$

$$\begin{matrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix}$$

$$\bar{x}_n \text{ nin olasılık yoğunluğu} \quad P_n = \frac{1}{\sqrt{2n}} e^{-1/2(\bar{x}_n - \theta)^2} \quad (3.9)$$

şeklindedir.

$\bar{x}_i$ ler bağımsızdır ve varyansı 1 olan bir normal dağılıma uyarlar. Bağımsız $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$ nin birlikte meydana gelmesi olasılığını maksimum yapan θ değeri örnek ortalamalarının ortalamasına eşittir.

Momentler Yönteminin, Maksimum Likelihood Yöntemine oranla daha sık kullanılmasının sebebi; tesadüfi değişkenlerin dağılımının bilinmesi gibi bir önşartı gerektirmemesidir.

3.3. Regresyon Yöntemi

Birçok istatistik çalışmada olduğu gibi regresyon yöntemi ile tahminde de, anakütle verilerinin tümü yerine anakütleden seçilen örnek verileriyle analiz yapılır⁴. Daha sonra elde edilen değerler anakütle için genellenerek tahminler yapılır. Anakütle ve örnek verileri kullanılarak gerçekleştirilen istatistiksel yöntemler arasında uygulama açısından fark yoktur, fakat anakütle parametreleri örnekleme teorisinden yararlanarak tahmin edilebilir.

Regresyon tahmininde öncelikle değişkenin ilişkisine göre uygun bir fonksiyon tipi seçilir. Daha sonra fonksiyonun parametreleri en küçük kareler yöntemiyle belirlenerek bağımsız değişkenin belli değerlerine karşılık bağımlı değişkenin alacağı değerler bu regresyon fonksiyonundan hesaplanır⁵.

Uygulamada kullanılan pekçok regresyon denklemi mevcuttur. Bu denklemlerden bazıları aşağıda gösterilmiştir:

- $Y = a + bx$
- $Y = a + bx + cx^2$
- $Y = a + bx + cx^2 + dx^3$
- $Y = a + bx + cx^2 + dx^3 + ex^4$
- $Y = \frac{1}{a + bx}$
- $Y = ab^x$ veya $\log Y = \log a + X \log b$
- $Y = ax^b$ veya $\log Y = \log a + b \log X$

Burada anlatılacak olan regresyon basit doğrusal regresyondur ve fonksiyon tipi $Y = a + bX$ dir.

Doğrusal regresyon tahmininde, N büyüklüğündeki anakütleden yola çıkılır. Bu anaküttelede X ve Y ile gösterilen iki değişkenin incelendiği düşünülür ve bu iki değişken arasında doğrusal bir ilişki olduğu varsayılır. X ve Y değişkenlerinden X bağımsız , Y ise bağımlı değişkendir. Bu iki değişken arasındaki basit doğrusal regresyon denklemi $Y = \alpha + \beta X$ ile gösterilir. Bu denklemde Y_i ler, X_i değişkenine karşılık gelen gözlenen değerlerdir ve Y'_i ler ise X_i lerin denklemde yerine konmasıyla elde edilir. Tahminlerde yapılan hatalar ise $Y_i - Y'_i = u_i$ ile ifade edilir. Denklem iki değişken arasındaki ilişkinin fonksiyonel şeklini göstermekle kalmaz, değişkenlerden birine ait değerler bilindiği takdirde diğerleri hakkında tahmin yapılmasını da

⁴ Orhunbilge,N.,Uygulamalı Regresyon ve Korelasyon Analizi, İstanbul: Avcıol Basımevi,1996, s.11

⁵ Çıngı,H., Örneklem Kuramı, Ankara: H. Ü. Fen Fakültesi Basımevi, 1990, s.171.

mümkün kılar. Regresyon ile tahmin yönteminde esas, tahmin hatalarının oluşturduğu teorik bölünmenin varyansının minimum olmasıdır.

Hata terimleri teorik bölünmesinin varyansını minimum kılmak, regresyon yönteminin esas amacıdır. Bu yüzden hata terimlerine ait varsayımları açıkça belirtmek gerekir. Bu varsayımlar aşağıdaki gibi sıralanabilir:

- Serbest değişken olan X ile hata terimleri arasında ilişki yoktur, yani birbirlerinden bağımsızdırlar. Bunu $E(u_i, X_i) = 0$ şeklinde göstermek mümkündür. Ancak bu varsayımın sağlanması, X in sabit bir değer olmasına bağlıdır. X bir tesadüfi değişken olduğunda bu varsayım sağlanmaz.

- Hata terimlerinin dağılması normaldir ve $E(u_i) = 0$ şeklinde gösterilir. Aksi halde normal dağılmama durumu ortaya çıkar.

- Herbir X_i değişkenine göre beliren Y_i alt anakütlelerinin varyansları sabittir yani $E(u_i^2) = S_u^2$ dir. Bu varsayım sağlanmadığı takdirde değişen varyanslılık durumu ortaya çıkar.

- Hata terimleri birbirinden bağımsızdır yani $E(u_i, u_j) = 0$ dir. Bu varsayım sağlanmadığı durumda ise özilişkinin varlığından sözedilir.

Yukarıdaki koşullar sağlandığı takdirde α ve β nın tahminleri EKK yöntemi ile bulunur ve bu yöntem en küçük varyansı verir. Daha açık bir şekilde ifade etmek gerekirse, N büyüklüğündeki anakütle parametreleri olan α ve β parametreleri EKK yöntemi ile hesaplandığı zaman bu noktaların X eksenine olan uzaklıklarının kareleri toplamı minimum olur. α ve β parametrelerinin tahminleri aşağıdaki gibi elde edilir:

$$\beta = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} \quad (3.10)$$

$$\alpha = \bar{Y} - \beta \bar{X} \quad (3.11)$$

Gözlenen değerlerin yani Y' lerin elde edilen $Y = \alpha + \beta X$ doğrusu etrafındaki dağılımı

$$S_{yx}^2 = \frac{\sum (Y - Y')^2}{N} \quad (3.12)$$

formülü ile hesaplanır. Bu değer korelasyon katsayısı ϕ 'ye göre de yazılabilir:

$$\phi^2 = 1 - \frac{S_{yx}^2}{\sigma_y^2} \quad (3.13)$$

$$\sigma_y^2 = \frac{\sum (Y - \bar{Y})^2}{N} \quad (3.14)$$

olduğu gözönünde bulundurulursa;

$$\phi = 0 \quad S_{yx}^2 = \sigma_y^2 \quad (3.15)$$

$$\phi = 1 \quad S_{yx}^2 = 0 \quad (3.16)$$

olur ve anakütle parametresi β ,

$$\beta = \phi \frac{S_y}{S_x} \quad (3.17)$$

şeklinde ifade edilir.

Anakütleyle ilişkin parametre tahminleri yapıldıktan sonra sıra N büyüklüğündeki bu anakütleden çekilen n birimli basit tesadüfi örnekleme örneğine ait a ve b değerlerinin hesabına gelmiştir. $y = a + bx$ doğru denkleminde bulunan bu değerlerin hesabı aşağıdaki gibidir:

$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad (3.18)$$

ve

$$a = \bar{y} - b\bar{x} \quad (3.19)$$

şeklinde hesaplanır.

Örnek değerleri olan a ve b yukarıdaki denklemlerden hesaplandığı taktirde,

- $E(b_{yx}) = \beta_{yx}$
- $E(a) = \alpha$
- $E(s_{yx}^2) = S_{yx}^2$

ve

$$s_{yx}^2 = \frac{\sum (y - y')^2}{n - 2} \quad (3.20)$$

dir.

b_{yx} regresyon parametrelerinin dağılması n-2 serbestlik dereceli student bölünmesi verir ve varyansı,

$$V(b_{yx}) = \frac{S_{yx}^2}{\sum (X - \bar{X})^2} = \frac{S_{yx}^2}{\sum X^2 - n\bar{X}^2} \quad (3.21)$$

ile hesaplanır. Anakütle değerleri bilinmediğinde yerine örnek değerleri konulursa varyansın tahmini değeri elde edilir:

$$\hat{\sigma}^2(b_{yx}) = \frac{s_{yx}^2}{\sum (x - \bar{x})^2} \quad (3.22)$$

Elde edilen verilere dayanarak β parametresinin geçerlilik hudutları

$$\hat{\beta}_{1,2} = b_{yx} \pm t\sigma_b \quad (3.23)$$

ile hesaplanır. $n \geq 30$ için normal alanlar tablosundan faydalanılır.

a değerleri teorik bölünmesine ait tahmini varyans ise

$$\hat{\sigma}^2(a) = s_{yx}^2 \frac{\sum x^2}{n \sum (x - \bar{x})^2} \quad (3.24)$$

dir ve α için geçerlilik hudutları ise aşağıdaki gibi bulunur.

$$\hat{\alpha}_{1,2} = a \pm t\sigma_a \quad (3.25)$$

Yukarıda geçerlilik hudutları için verilen fomüllerde yeralan t değeri, student tablosundan alınacaktır.

Anakütle parametresi olan ortalamanın tahmininde doğru orijinden geçmediği için tahmin;

$$\hat{Y} = \bar{y} + b(\bar{X} - \bar{x}) \quad (3.26)$$

den hesaplanır.

b değeri yerine bu kez

$$b = r \frac{s_y}{s_x} \quad (3.27)$$

kullanılırsa denklem,

$$\hat{Y} = \bar{y} + r \frac{s_y}{s_x} (\bar{X} - \bar{x}) \quad (3.28)$$

şeklını alır.

Tahminlerden oluşan teorik bölünmenin varyansı ise

$$V(\bar{y}) = \frac{N-n}{N} \frac{s_y^2}{n} (1-r^2) \quad , \quad \frac{n}{N} > 0.05 \quad (3.29)$$

$$V(\bar{y}) = \frac{s_y^2}{n} (1-r^2) \quad , \quad \frac{n}{N} < 0.05 \quad (3.30)$$

ile hesaplanır.

Görüldüğü gibi yukarıdaki ifadede yeralan ilk terim Basit Tesadüfi Örneklem örneği ele alınarak elde edilen tahminler bölünmesinin varyansını verir. İkinci terim olan $(1-r^2)$ ise regresyon yöntemi ile tahminden doğan çarpandır ve daima 1 den küçüktür ,dolayısı ile bu varyans Basit Tesadüfi Örneklem örneğine ait varyanstan daha küçük çıkar.

3.4. Nisbet Yöntemi

N birimli bir anakütleden Basit Tesadüfi Örneklemeye yöntemine dayanarak n sayıda birim içeren bir örnek çekilir ve bu örnekteki her birim X ve Y vasıflarına göre incelenirse sonuçta n tane $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ çifti oluşur. x_i lerin toplamı ve ortalaması $\bar{x} = \frac{\sum x_i}{n}$, y_i lerin toplamı ve ortalaması $\bar{y} = \frac{\sum y_i}{n}$ hesaplanır. İncelenen X ve Y vasıfları arasında doğrusal bir ilişki olduğu ve X vasfına ait anakütle toplam ve ortalamasının belli olduğu varsayılırsa bu durumda

$$y = bx \quad (3.31)$$

$$\bar{y} = b\bar{x} \quad (3.32)$$

$$\bar{Y} = b\bar{X} \quad (3.33)$$

den yola çıkarak

$$\frac{\bar{X}}{\bar{x}} = \frac{\bar{Y}}{\bar{y}} \quad \text{veya} \quad \bar{Y} = \bar{X} \frac{\bar{y}}{\bar{x}} \quad (3.34)$$

yazılabilir. Tahmini ise aşağıdaki gibi elde edilir:

$$\hat{Y} = \bar{X} \frac{\bar{y}}{\bar{x}} = \bar{X} \frac{y}{x} \quad (3.35)$$

Buradan Y nin tahmini

$$Y = N\hat{Y} = N\bar{X} \frac{\bar{y}}{\bar{x}} = N\bar{X} \frac{y}{x} \quad (3.36)$$

şeklinde ifade edilir.

Nisbet usulüne göre tahmin yönteminin başarılı olabilmesi için şu şartların sağlanması gerekir:

- $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ noktaları arasından geçirilecek doğrunun orijinden veya orijine çok yakın bir yerden geçmesi gereklidir.

- X_i ler büyüdükçe Y_i lerin varyansı da büyüyorsa nisbet usulü ile tahminin doğruluğu da artar.



4. ÖRNEKLEME YÖNTEMLERİ

İyi bir örnek, anakütleye en uygun örnekleme yönteminin belirlenmesinden sonra örneklem genişliğinin saptanmasıyla elde edilebilir. Uygun örnekleme yönteminin belirlenmesi parametreye ilişkin tahminler bölünmesi varyansının minimum kılınmasıyla mümkündür. Bu nedenle uygun örnekleme yönteminin seçimi istatistikte önemli bir yer tutar.

Bu bölümde belli başlı örnekleme yöntemlerinden bahsedilecektir. Bunlar sırasıyla; Basit Tesadüfi Örnekleme, Zümrelere Göre Örnekleme, Sistemantik Örnekleme ve Kümelere Göre Örneklemedir.

4.1. Basit Tesadüfi Örnekleme

Uygulamada başvurulabilecek birçok örnekleme tekniği mevcuttur. Ancak Basit Tesadüfi Örnekleme , diğer örnekleme teknikleri içinde en yaygın olarak bilinen örnekleme türüdür. Bu yüzden öncelikle bu yöntemin açıklanması uygun görülmüştür.

Basit Tesadüfi Örnekleme kısaca, N büyüklüğündeki bir anakütleden her biri n birimden oluşan C_n^N sayıda örnek çekme yöntemidir⁶. Bu örneklerin hepsi eşit seçilme şansına sahiptir. Uygulamada n birimli örneklerin oluşumu, birimleri teker teker çekmek suretiyle oluşturulur ve her birimin örneğe girme olasılığı n/N olduğundan çekim birer birer yapılır. Önce anakütledeki birimler 1 den N 'e kadar numaralanır daha sonra bunların arasından n tanesi kura veya tesadüfi sayılar tablosu yardımıyla çekilir. Birim sayısı az ise kura yöntemi uygulanabilir. Ancak birim sayısı

⁶ Cochran ,W.G., Sampling Techniques, New York: John Wiley & Sons, 1977, s.18.

arttıkça tesadüfi sayılar tablosundan yararlanmak mecburi hale gelir. Tesadüfi sayılar tabloları çok sayıdaki rakamların ardarda sıralanmasıyla elde edilmiş olup, birbirini izleyen rakamlar arasında hiçbir ilişki yoktur. Bu tablolar günümüzde bilgisayar vasıtasıyla yaratılmaktadır.

n büyüklüğündeki bu örneklerin herbirinin eşit çekilme şansına sahip olduğunu kanıtlamak oldukça basittir. Tek bir örnek, yani n sayıda birimden oluşan tek bir seri düşünüldüğünde, n tane birimden birini çekme olasılığı n/N dir. İkinci çekme işleminde, geriye kalan $(n-1)$ birimden birini çekme olasılığı $(n-1)/(N-1)$ dir ve bu şekilde n tane birim çekilene kadar işlem devam eder. Çekilen bu örnekteki her birim örnekte yalnız ve yalnız bir kez yer almakta yani çekim iadesiz yapılmaktadır. Bu kural C_n^N sayıdaki bütün n birimli örnekler için geçerlidir. Böylece n birimden oluşan örneğin hepsinin birden çekilme olasılığı

$$\frac{n}{N} \cdot \frac{(n-1)}{(N-1)} \cdot \frac{(n-2)}{(N-2)} \cdots \frac{1}{(N-n+1)} = \frac{n!(N-n)!}{N!} = \frac{1}{C_n^N} \quad (4.1)$$

olur⁷ ve anakütleye dahil her birimin çekilecek n birimli örneğe girme olasılığı n/N dir. Ayrıca bu oran tüm birimler için aynıdır. Yukarıda da belirtildiği gibi çekilen her birim anakütleden çıkarıldığı için bu yöntem iadesiz basit tesadüfi örnekleme adını alır.

Şimdi büyük harfler anakütleyi, küçük harfler ise örneği göstermek üzere işlemlerde kullanılacak notasyon verilecektir:

Anakütleyi meydana getiren N tane birimin herbiri için elde edilen değerler $X_1, X_2, \dots, X_N$ ile gösterilir. Örnekte bu birimlere karşılık gelen değerler ise $x_1, x_2, \dots, x_n$ şeklinde ifade edilir. Genel bir notasyon aşağıdaki gibi verilebilir:

⁷ Cochran, W.G., a.g.e., s.18.

Anakütle

$$\text{Top.} \quad X = \sum_{i=1}^N X_i = X_1 + X_2 + \dots + X_N$$

$$\text{Ort.} \quad \bar{X} = \frac{X_1 + X_2 + \dots + X_N}{N} = \frac{\sum_{i=1}^N X_i}{N}$$

$$\text{Dağılım} \quad V(X) = \sigma^2 = \frac{\sum (X_i - \bar{X})^2}{N}$$
$$S^2 = \frac{\sum (X_i - \bar{X})^2}{N - 1}$$

Örnek

$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$$

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x_i}{n}$$

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

Anakütle ortalaması bilinmediği takdirde, bu değer anakütle içinden çekilen n birimli örnek ortalamasından hareketle bir sayı veya bu sayıyı kapsayan bir aralık şeklinde tahmin edilir. Anakütle ortalaması $\bar{X}$, örnek ortalaması olan $\bar{x}$ ile tahmin edildiği zaman tek bir değer ile tahmin söz konusu olur ve bu tahmin nokta tahmin adını alır. Bu ifade aşağıdaki gibi gösterilir:

$$\hat{\bar{X}} = \bar{x} = \text{örnek ortalaması} \quad (4.2)$$

Anakütle toplamı ise aşağıdaki gibi ifade edilir:

$$\hat{X} = N\bar{x} = N \sum x_i / n \quad (4.3)$$

4.1.1. Örnek Ortalamaları Teorik Bölünmesi

Bir anakütleden çekilebilecek tüm n birimli örneklerin çekildiğini ve bu örnekler için ortalamaların hesaplandığını ($\bar{x}_i$) düşünelim. Bu ortalamaların oluşturduğu teorik bölünme serisinde herbir örneğin ortalaması anakütle ortalamasının bir tahminidir. Bir başka ifade ile örnek ortalamalarının beklenen değeri anakütle ortalamasına eşittir kısaca, örnek ortalaması $\bar{x}$, anakütle ortalaması $\bar{X}$ nın yansız bir tahminidir ve aşağıdaki gibi ifade edilir.

$$E(\bar{x}) = \bar{\bar{x}} = \bar{X} \quad , \quad \bar{x}_i = \hat{\bar{X}} \quad (4.4)$$

Örnek ortalamalarından oluşan teorik bölünmenin varyansı ise;

$$V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \frac{N-n}{N-1} \quad , \quad \frac{n}{N} > 0.05 \quad (4.5)$$

$$V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \quad , \quad \frac{n}{N} < 0.05 \quad (4.6)$$

ile hesaplanır.

Anakütle varyansı σ^2 bilinmiyorsa

$$V(\bar{x}) = \frac{\sigma^2}{n} \frac{N-n}{N-1} \quad (4.7)$$

formülünde σ^2 yerine örnek varyansı olan σ^2 kullanılamaz. Çünkü örnek varyansı anakütle varyansının yansız bir tahmini değildir. Bu durumda $S^2 = \sigma^2 \frac{N}{N-1}$ özelliğinden yararlanılır. Ancak çoğu zaman S^2 bilinmediği için yerine onun yansız bir tahmini olan s^2 kullanılır, böylece örnek ortalamaları teorik bölünmesinin varyansının tahmini;

$$V(\bar{x}) = \hat{\sigma}_{\bar{x}}^2 = \frac{s^2}{n} \frac{N-n}{N} \quad , \quad \frac{n}{N} > 0.05 \quad (4.8)$$

veya

$$V(\bar{x}) = \hat{\sigma}_{\bar{x}}^2 = \frac{s^2}{n} \quad , \quad \frac{n}{N} < 0.05 \quad (4.9)$$

şeklinde hesaplanır.

N büyüklüğündeki bir anakütleden çekilmesi olası n birimli tüm C_n^N sayıdaki örneklerden biri çekildiğinde bu örneğin ortalamasının geçerlilik hudutları ise aşağıdaki gibi bulunur:

$$\bar{x}_{1,2} = \bar{X} \pm t\sigma_{\bar{x}} \quad (4.10)$$

Ancak burada bilinmesi gereken iki değer sözkonusudur. Bunlar anakütle ortalaması ($\bar{X}$) ve örnek ortalamaları teorik bölünmesinin

varyansı σ_x dır. $\bar{X}$ bilinmediğinde bu değer örnek ortalaması $\bar{x}$ ya bağlı olarak tahmin edilir ki bu ifade (4.2) nolu formül ile açıklanmıştır. Örnek ortalamaları teorik bölünmesinin varyansının hesabı da yukarıda verilmiştir. Böylece bilinmeyen anakütle parametreleri yerine yansız tahminleri konularak hesap yapılır.

Anakütle ortalamasının geçerlilik hudutlarının hesabı ise aşağıdaki gibidir:

$$\hat{\bar{X}}_{1,2} = \bar{x} \pm t\sigma_{\bar{x}} \quad (4.11)$$

Anakütle ortalamasına ait tahminin güvenirliliği örnek ortalamaları teorik bölünmesinin dağılmasının küçük çıkmasına bağlıdır. Bu da standart hatanın küçük çıkması demektir. Yani

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}, \quad \frac{n}{N} < 0.05 \quad (4.12)$$

in küçük çıkması

- n sabit iken σ nın küçülmesine
- σ sabit iken n in büyümesine

bağlıdır. n büyüdüğü zaman maliyette ve sistematik hatalarda artış meydana gelir. Bu yüzden n in büyümesi yerine σ nın küçültülmesi tercih edilir. Ancak σ , anakütle standart sapması olduğu ve genelde veri durumunda bulunduğu için standart hatanın küçülmesi için ya tahmin yönteminin veya örnekleme yönteminin değiştirilmesi gerekmektedir. Bu yüzden Basit Tesadüfi Örnekleme yöntemi kullanıldığında ya Momentler Yöntemi yerine Regresyon Yöntemi ile tahmin yapılır veya örnekleme yöntemi değiştirilerek Zümrelere Göre Örnekleme Yöntemi kullanılır.

4.1.2. Örnek Varyansları Teorik Bölünmesi

Örnek varyansları teorik bölünmesinde de, yine N büyüklüğündeki anakütle ele alınır ve bu anakütleden n sayıda birim içeren C_n^N tane örnek çekilir. Daha sonra bu örneklerin herbiri için örnek varyansları σ_i^2 ler hesaplanır.

Örnek ortalamaları teorik bölünmesinde geçerli olan özelliklerden birincisi burada geçerli değildir. Yani, örnek varyanslarının oluşturduğu teorik bölünmenin beklenen değeri anakütle varyansına eşit değildir. Örnek varyansları teorik bölünmesinin beklenen değeri

$$E(\sigma^2) = \frac{\sum \sigma_i^2}{\binom{N}{n}} \quad (4.13)$$

$$E(\sigma^2) = \sigma^2 \frac{n-1}{n} \frac{N}{N-1} \quad (4.14)$$

şeklinde ifade edilir. Fakat N çok büyük olduğunda N ve N-1 arasındaki fark önemsiz bir hale gelecek ve böylece yukarıdaki ifade

$$E(\sigma^2) = \sigma^2 \frac{n-1}{n} \quad (4.15)$$

şeklini olacaktır. Yukarıdaki ifadeden de anlaşılacağı gibi $E(\sigma^2) < \sigma^2$ dir ve böylece örnek varyansları teorik bölünmesinin beklenen değerinin anakütle varyansından büyük olduğu görülür.

Diğer taraftan $S^2 = \sigma^2 \frac{N}{N-1}$ ifadesinde , N in büyük değerleri için $S^2 = \sigma^2$ yazılabilir. Örnek s^2 ler teorik bölünmesinin beklenen değeri anakütle S^2 sine eşit olduğundan dolayı $E(s^2) = \sigma^2$ yazılabilir. Yani anakütle varyansının tahmini olarak örnek s^2 si kullanılabilir.

Örnek varyanslarından oluşan teorik bölünmenin varyansı ise anakütle dağılımının normal olup olmamasına göre iki şekilde ortaya çıkar:

- Anakütle dağılımı normal ise

$$V(s^2) = \sigma_{s^2}^2 = \frac{2\sigma^4}{n-1} \quad (4.16)$$

- Anakütle dağılımı normal değilse

$$V(s^2) = \sigma_{s^2}^2 = \frac{2^4}{n-1} + \frac{K_4}{n} \quad (4.17)$$

$$K_4 = E(x_i - \bar{x})^4 - 3\sigma^4 \quad (4.18)$$

şeklinde hesaplanır.

Anakütle varyansının geçerlilik hudutları ise

$$P\left\{\frac{(n-1)s^2}{\chi_1^2} > \sigma > \frac{(n-1)s^2}{\chi_2^2}\right\} = 1 - \alpha \quad (4.19)$$

şeklinde ifade edilir. χ_1^2 için $\alpha/2$ tablo değeri, χ_2^2 için $1-\alpha/2$ tablo değeri alınır. $n \geq 30$ için χ^2 bölünmesi, standart normal bölünmeye yaklaşır.

4.1.3. Örnek Nisbetleri Teorik Bölünmesi

Anakütlede bulunan bütün birimlerin sayısı N ve belli bir özelliği gösterenlerin sayısı da A ile gösterildiğinde anakütle nisbeti ;

$$P = \frac{A}{N} \quad (4.20)$$

şeklinde ifade edilir. Anakütlede nisbeti, N birimli anakütle içinden n birim içeren C_n^N tane örneğin nisbetinden yola çıkarak bulunur. Yani kısaca anakütle nisbeti, örnek nisbetlerinden oluşan teorik bölünmenin beklenen değerine eşittir ve aşağıdaki gibi ifade edilir:

$$E(p) = P \quad (4.21)$$

Örnek nisbetlerinden oluşan teorik bölünmenin varyansı aşağıdaki gibi hesaplanır:

$$V(p) = \frac{PQ}{n} \frac{N-n}{N-1} \quad , \quad \frac{n}{N} > 0.05 \quad (4.22)$$

$$V(p) = \frac{PQ}{n} \quad , \quad \frac{n}{N} > 0.05 \quad (4.23)$$

dir. Anakütle parametreleri olan P ve Q (P+Q=1) bilinmediği takdirde yerine tahmini değerleri olan p ve q yazılırsa, örnek nisbetleri teorik bölünmesinin varyansının tahmini elde edilir:

$$\hat{\sigma}^2(p) = \frac{pq}{n-1} \quad , \quad \frac{n}{N} < 0.05 \quad (4.24)$$

Örnek nisbetleri teorik bölünmesinin şeklini açıklarken p_i nin bir binom dağılıma değişkeni olduğunu söylemek gerekir. p_i ler bölünmesi çekim iadeli ise binom, iadesiz ise hipergeometrik dağılıma uyar. Anakütle nisbetinin belli olup olmamasına göre bazı özel durumlar mevcuttur ve özetle aşağıdaki gibi verilir:

- P belli ve 0.5 ise;

Küçük örnekler için bile örnek nisbetleri teorik bölünmesi normal bölünmeye yaklaşır. Uygulamada $n \geq 20$ için bu yaklaşım uygun görülmekte, bazen $20 > n > 9$ değerleri için binom bölünme yapılan süreklilik düzeltmesi ile normal bölünmeye dönüştürülebilmektedir.

$$t = \frac{p - P}{\sigma_p} \quad (4.25)$$

$$\sigma_p = \sqrt{\frac{PQ}{n}} \quad (4.26)$$

de süreklilik düzeltmesi yapılarak,

$$t = \frac{p - \frac{1}{2n} - P}{\sigma_p} \quad (4.27)$$

veya

$$t = \frac{|p - P| - \frac{1}{2n}}{\sigma_p} \quad (4.28)$$

şeklinde ifade edilir.

Anakütle nisbetinin geçerlilik hudutları ise

$$\hat{P}_{1,2} = p \pm t \sqrt{\frac{PQ}{n}} \quad (4.29)$$

ile hesaplanır.

- P belli fakat 0.5 değilse

Bu durumda n'in belli bir değeri için anakütle nisbeti P, 0.50'ye mümkün olduğu kadar yaklaştıkça örnek nisbetleri teorik bölünmesi normale yaklaşır.

- P belli değilse

Uygulamada genellikle anakütle nisbetinin bilinmediği durum ile karşılaştırılır. Bu durumda $E(p) = P$ özelliğinden faydalanılarak $p_i = \hat{P}$ olduğu gözönünde bulundurulur ve anakütle nisbeti aşağıdaki şekilde tahmin edilir:

$$\begin{aligned} \hat{P}_{1,2} &= p \pm t \sigma_p \\ &= p \pm t \sqrt{\frac{PQ}{n}} \\ &= p \pm \left\{ t \sqrt{\frac{PQ}{n}} \sqrt{\frac{N-n}{N-1}} \right\} \end{aligned} \quad (4.30)$$

ile tahmin edilir.

Anakütle nisbeti bilinmediği zaman yerine tahmini değeri konulur ve ifade

$$\hat{P}_{1,2} = p \pm t \sqrt{\frac{pq}{n-1}} \quad (4.31)$$

şeklini alır.

4.2. Zümrelere Göre Örneklem

Zümrelere Göre Örneklem yönteminin esası anakütleyi, değerleri birbirine yakın olan birimleri biraraya toplayarak varyansları küçük ve zümre adı verilen homojen gruplar şeklinde yeniden oluşturmaktır. Daha sonra da bu homojen grupların herbirini örneklemeye tabi tutarak anakütle parametrelerine ilişkin tahminler yapılır. Zümrelere Göre Örneklemde amaç tahminlerin güvenilirliğini arttırmaktır⁸. Bu da ancak grupların kendi içlerinde homojen yani ortalamaları birbirinden çok farklı olacak şekilde oluşturulmasına bağlıdır. Homojenlik arttıkça zümreden örneğe alınacak birimlerin sayısı da azalacaktır.

Zümrelere Göre Örneklem yöntemi ile ilgili; anakütle ortalamasının, varyansının ve nisbetinin tahminlerine ilişkin formüller ve açıklamalar daha sonraki bölümlerde verilecektir. Şimdilik sadece yöntemin içeriğinden kısaca bahsetmekle yetinilecektir.

4.3. Kümelere Göre Örneklem

Kümelere Göre Örneklem, Çok Kademeli Örneklemenin özel bir şeklidir. Bu örneklem tekniğinde seçim; asıl birimler arasından değil, mensup oldukları kümeler arasından yapılır.

Daha önce de ifade edildiği gibi Zümrelere Göre Örneklem yönteminde, anakütle zümrelere ayrılıyor sonra da herbir zümreye örneklem uygulanarak işlemler yapılıyordu. Kümelere Göre Örneklem Yönteminde ise anakütle birtakım kümelere ayrılıyor, bu kümeler üzerinden örneklem yapılıyor ve incelenen kümelerin tamamı işleme tabi tutuluyor.

⁸ Hahn, G.J., Meeker, W.Q., Statistical Intervals, New York: John Wiley & Sons, 1991, s. 12.

Kümelere Göre Örneklem Yönteminde, anakütle hiçbir birim açıkta kalmayacak şekilde kümelere bölünür ve bu kümeler arasından tesadüfi seçimle iki veya daha fazlası örnek olarak alınır. Zümrelere Göre Örneklemde oluşturulan zümrelerin ortalamaları birbirinden oldukça uzak olacak şekilde teşkil ediliyor ve bu zümreler içinden birer örnek çekiliyordu. Ancak birimlerin birbirinden uzak olması, yani zümrelerin kendi içlerinde homojen fakat aralarında heterojen olması, gözlem ve ulaşım masraflarını artırıyor ayrıca araştırmanın tamamlanması için büyük harcamalara sebebiyet veriyordu. Fakat, araştırma çerçevesinin belirlenemediği veya ulaşma maliyetinin çok yüksek olduğu hallerde Kümelere Göre Örneklem yönteminden yararlanılabilir⁹. Kümelere Göre Örneklem yönteminde , oluşturulan kümeler ortalamaları birbirine yakın olacak şekilde belirlenir, örneğin oluşturulan 8 küme yerine onu temsil edebileceği düşünülen iki veya daha fazla küme incelenir ve böylece araştırmanın maliyeti azaltılabilir. Ancak, kümelere ayrılarak yapılan bu örnek çekme işleminde bazı kümelere bulunan birimler birçok yönden birbirlerine benzerlik gösterebilirler. Bu tür kümeleri inceleyerek yapılan anakütle tahminleri doğruluktan uzaktır ve örneklem hatası da büyüktür. Bu nedenle, kümelere göre örnekleme yaparak yapılan anakütle tahminlerinin temsili yeteneği, basit tesadüfi örnekleme göre daha zayıftır. Örneklem hatasını azaltmanın yolu ise, örneği çok sayıda kümeden çekilecek birimlerden oluşturmaktır.

Çalışmada esas, Zümrelere Göre Örneklem yöntemine ağırlık vermek olduğu için bu bölümde sadece Tek Kademeli Örneklem veya Kümelere Göre Örneklem yönteminden bahsedilecektir.

Oluşturulacak kümelerin büyüklüklerinin farklı olması araştırmacıya zorluk çıkarabilir. Bu yüzden genellikle kümeler eşit sayıda birim içerecek şekilde oluşturulurlar.

⁹ İdil,O., Örneklem Teorisi ve İşletme Yön Uyg., İstanbul: Fatih Yayınevi Mat., 1980, s. 213

Bu bölümde incelenecek olan Tek Kademeli Örneklemeye , küme birim sayısı eşit olduğunda aşağıdaki aşamalardan geçirilerek yapılır:

- Anakütle her biri B sayıda birim içeren A tane kümeye ayrılır.
- Oluşturulan bu kümelerden a tanesi Basit Tesadüfi Örneklemeye yöntemiyle seçilir.
- Seçilen bu a tane kümeye tamsayım uygulanır.

Yukarıdaki açıklamalara dayanarak anakütle ve örnek büyüklükleri belirlenebilir. Anakütle büyüklüğü; B sayıda birim içeren A tane küme olduğu için $N=A \times B$ şeklinde ifade edilir. Örnek büyüklüğü belirlenirken A tane küme içinde a tanesi çekildiği için bu kez B sayıda birim içeren a tane küme gözönünde bulundurulur. Bunlara ilave olarak anakütleye dahil her birimin çekilecek örneğe girme olasılığı

$$\frac{a}{A} = \frac{a}{A} \frac{B}{B} = \frac{n}{N} \quad (4.32)$$

şeklinde hesaplanır.

Kümelere Göre Örneklemeye n sayıda birimden oluşan örneğin ortalaması anakütle ortalamasının yansız bir tahminidir ve aşağıdaki şekilde gösterilir:

$$E(\bar{x}) = \bar{X} \quad , \quad \bar{x} = \hat{\bar{X}} \quad (4.33)$$

Örnek ortalaması yani anakütle ortalamasının tahmini ise

$$\hat{\bar{X}} = \bar{x} = \frac{1}{a} \sum_{i=1}^a \bar{x}_i \quad (4.34)$$

formülü ile bulunur. Burada örnek ortalaması yerine

$$\bar{x}_i = \frac{1}{B} \sum x \quad (4.35)$$

yazılırsa anakütle ortalamasının tahmini bu kez;

$$\hat{\bar{X}} = \frac{1}{a \cdot B} \sum x = \frac{1}{n} \sum x \quad (4.36)$$

şeklinde elde edilir.

Kümelere Göre Örneklemede tahminler bölünmesinin varyansı ise aşağıdaki gibidir:

$$V(\bar{x}) = \frac{A-a}{A} \frac{S_a^2}{a} \quad (4.37)$$

Burada;

$$S_a^2 = \frac{1}{A} \sum_{i=1}^A (\bar{x}_i - \bar{X})^2 \quad (4.38)$$

ile bulunur.

Küme büyüklüğü eşit olmadığı zaman ise;

- Anakütle ortalamasının tahmini,

$$\hat{\bar{X}} = \bar{x} = \frac{\sum B_i \bar{x}_i}{\sum B_i} \quad (4.39)$$

ve

- Tahminler bölünmesinin varyansı

$$V(\bar{x}) = \frac{A-a}{A} \frac{s_a^{2''}}{a} \quad (4.40)$$

$$s_a^{2''} = \frac{1}{a-1} \sum \frac{B_i^2}{B^2} (\bar{x}_i - \bar{x})^2 \quad (4.41)$$

şeklinde hesaplanır¹⁰.

Konunun başında da ifade edildiği gibi, Kümelere Göre Örneklemenin esas amacı ortalamaları birbirine yakın kümelerden iki veya daha fazlasını inceleyerek maliyeti azaltmaktır. Ancak incelenecek küme sayısı az olduğunda varyansın büyümesi de kaçınılmazdır. Kademe sayısı artırılarak bu sorun aşılmaya çalışılır.

¹⁰ Yoğurtçugil, M.K., a.g.e.,s.

4.4. Sistematik Örneklem

Sistematik örneklem yöntemi bazı yönleriyle Basit Tesadüfi Örneklemeye benzerlik gösterir. Sistematik Örneklemde, anakütle yine 1'den N'e kadar numaralanır. Numaralanan bu anakütleden n büyüklüğünde bir örnek oluşturabilmek için ilk k birim içinden rasgele bir birim çekilir. Eleman sayısı fazla olan anakütlelerde her birimin tek tek incelenmesinin masraflı olduğu durumlarda kullanılan sistematik örneklemde, çekilen rasgele birim başlangıç noktası olarak kabul edilir.

Sistematik örneklem yöntemi kısa sürede sonuç veren ve sonucun elde edilebilmesi için fazla bir masraf yapılmasına gerek duymayan bir yöntemdir. Sistematik örneklem yöntemine göre bir örneğin düzenlenebilmesi için örneğin alınacağı anakütle tesadüfi olarak düzenlenmiş ve tesadüfi olarak ele alınan anakütlenin elemanları önem derecelerine göre sıralanmış olmalıdır. Tabiki, bir örneğin düzenlenebilmesi için de, istenilen özelliklere sahip olan bu anakütleyle ait bir listenin mevcut olması gerekmektedir.

Sistematik Örneklemde; öncelikle anakütle 1 den N'e kadar numaralanır. Daha sonra $N/n=k$ büyütme faktörü hesaplanır. Bu oran örneklem oranı olan n/N in tersidir. Bu işlemler tamamlandıktan sonra ilk k tane birim arasından bir tanesi tesadüfi olarak çekilir. Şunu da ilave etmek gerekir ki bu yöntemde k, tamsayı olacak şekilde önceden tesbit edilir. Sonra buna bağlı olarak n belirlenir. Bu ilk k tane tamsayı arasından tesadüfen çekilen ilk sayıya, yani başlangıç noktası olarak kabul edilen sayıya a denilirse bundan sonraki n birimli örneğe sırası ile a, a+k, a+2k,.....,a+(n-1)k terimleri girecektir. Örnek vermek gerekirse, N=800 ve n=80 ise $k=800/80=10$ dir. İlk 10 birim arasından tesadüfen 8. birim başlangıç noktası olarak belirlenirse bundan sonra örneğe girecek birimlerin numarası sıra ile 8,18, 28,.....,398 olacaktır.

İlk terimin, yani başlangıç noktasının seçiminde ise a'nın tesadüfen seçilmesi yerine kullanılan bir diğer usulde k'nın tek ya da çift olmasına göre belirlenir. a; k çift ise k/2. terim, k tek ise (k+1)/2. terim olarak belirlenir. k tamsayı değil ise, örnek büyüklüğünde k'yı tamsayı çıkaracak şekilde ayarlama yapılır. Her bir örneğin çekilme olasılığı ise 1/k'dır. Çok büyük anakütller ile çalışıldığında k'nın tamsayı ya da tamamlanmamış şeklinin alınması fazla önemli değildir¹¹.

Sistematiik örneklemede, örnek ortalaması anakütle ortalamasının yansız bir tahminidir ve aşağıdaki şekilde gösterilir:

$$E(\bar{x}) = \bar{X} \quad , \quad \bar{x}_i = \hat{\bar{X}} \quad (4.42)$$

dır.

Anakütle ortalaması ise k sayıdaki örneğin ortalaması şeklinde ifade edilir:

$$\bar{X} = \frac{1}{k} \sum \bar{x}_i \quad (4.43)$$

Örnek ortalamalarından oluşan teorik bölünmenin varyansı ise;

$$V(\bar{x}) = \frac{\sum (\bar{x}_i - \bar{x})^2}{k} \quad (4.44)$$

şeklinde ifade edilir.

Sistematiik örneklemenin bir sakıncası, örnekleme aralığının iyi belirlenememesi durumunda her seferinde belirli bir özelliğe sahip olan birimler gereksiz yere örneğe dahil edildiğinde ortaya çıkar ve sistematiik bir hata oluşur. Bu sorunu ortadan kaldırmak için işlemler her seferinde tesadüfi olarak belirlenen yeni bir başlangıç noktası esas alınarak yapılır.

Sistematiik Örneklemenin tercih edileceği koşulları genellemek oldukça zordur. Etkin bir şekilde kullanabilmek için anakütlenin yapısını iyi

¹¹ Çingil.H., a.g.e., s. 196.

bilmek gerekir. Mesela, anakütlenin birimleri periyodik bir şekilde sıralanmışsa sistematik örneklemenin duyarlılığı tehlikeye düşebilir¹².



¹² Çıngı, H., a.g.e., 208

5. ZÜMRELERE GÖRE ÖRNEKLEME

Anakütle ortalaması veya toplamına ait bir tahminin doğruluğunun sadece, örnek büyüklüğü ve örnekleme oranına değil, aynı zamanda anakütle birimleri arasındaki değişkenliğe veya heterojenliğe de bağlı olduğu söylenebilir¹³. Anakütle ortalamasını çok yüksek bir doğruluk çerçevesinde tahmin etmenin bir yolu da; örnek büyüklüğünü arttırmak yerine anakütleyi, anakütlenin tamamından daha homojen olan alt gruplara bölmek ve grupların herbirinden büyüklükleri daha önceden belirtilmiş olan bir tesadüfi örnek çekmektir. Anakütlenin bölünen bu gruplarına zümre ve anakütleyi zümrelere bölme işleminin tamamına da zümrelere göre tesadüfi örnekleme denilir. Zümrelere Göre Örneklemeyi daha düzgün bir şekilde ‘ anakütle adı verilen yığını değerleri birbirine çok yakın olan birimleri biraraya toplamak suretiyle, varyansları küçük ve zümre adını alan homojen gruplar veya alt anakütleler şekline dönüştürmek ve zümrelerden herbirini ayrı birer bağımsız anakütle gibi düşünerek Basit Tesadüfi Örnekleme tabi tutmak’ şeklinde tanımlanarak mümkündür.

Basit Tesadüfi Örneklemede, anakütlenin bütün özelliklerinin örnekte de bulunacağı garanti değildir. Zümrelere Göre Örneklemede ise, anakütlenin farklı özelliklerini gösteren bir örnek çekmek mümkündür. Bu yüzden Zümrelere Göre Örnekleme örnek büyüklüğüne ilişkin bilginin varolduğu ve herbir zümreden bir örnek çekebilmek için gerekli olan çerçevenin mevcut olduğu araştırmalarda sıkça kullanılan bir yöntemdir.

¹³Sukhatme, P.V., Sampling Theory of Surveys with Applications, New York: The Iowa State Collage, 1958, s. 105

Anakütleyi zümrelere ayırmanın başlıca nedeni, anakütle dağılıma ölçüsü olan varyansı azaltmaktır. Bunun iki yolu vardır; birincisi örnek büyüklüğünü arttırmak ve böylece varyansı küçültmektir, ikincisi ise anakütleyi değerleri birbirine çok yakın olan homojen alt anakütlelere bölerek varyansları küçük homojen gruplar oluşturmak ve bu grupların herbirine Basit Tesadüfi Örneklemeyi uygulamaktır. Daha sonra da hesaplanan örnek değerleri birleştirilerek anakütle parametresi tahmin edilir. Zümrelere Göre Örneklemenin gerekliliğini bir örnekle açıklamak mümkündür: mesela; hane başına ortalama geliri tahmin etmek için, haneleri gelirlerine göre iki veya daha fazla gruba ayırmak ve tahmini bu gruplara göre yapmak daha sağlıklı sonuçlar verir. Anakütlenin tamamı ile karşılaştırıldığında, zümrelere ayrılan haneler daha homojendir. Böylece, zümrelere ayrılmış bir örneğe dayanarak yapılan hane başına gelirin tahmini değeri, aynı büyüklükteki basit tesadüfi bir örneğe dayanarak yapılan tahminden daha fazla güvenilirliğe sahiptir.

Zümrelere Göre Örneklemenin adımları aşağıdaki gibi sıralanabilir:

- Örneklem birimlerinin oluşturduğu anakütlenin tamamı zümre denilen alt anakütlelere bölünür.
- Herbir zümreden bir örnek çekilir.
- Herbir zümreden çekilen örnek için zümre ortalaması hesaplanır. Bu zümre ortalamaları ağırlıklandırılarak anakütlenin tamamına ilişkin tahmini yapmak üzere uygun bir hale getirilir.
- Anakütle varyansı ise, küçük varyanslı zümrelerin varyanslarının tartılı varyansı ile zümre ortalamalarının varyansının toplamı olur¹⁴.

Zümrelere Göre Örnekleme yönteminde, N birimli anakütle sırası ile $N_1, N_2, \dots, N_h, \dots, N_k$ birimli alt anakütlelere bölünür. Herbir birim sadece bir

¹⁴ Kish, L., Survey Sampling, New York: John Wiley & Sons, 1965, s.75.

örnekte bulunacağı için zümre büyüklüklerinin toplamı anakütle mevcuduna eşittir ve aşağıdaki gibi ifade edilir:

$$N = \sum_{i=1}^k N_h \quad (5.1)$$

Homojen gruplardan yani varyansları küçük zümrelerden Basit Tesadüfi Örnekleme yöntemine göre birer örnek çekilir ve bu örnek büyüklükleri de $n_1, n_2, \dots, n_h, \dots, n_k$ ile gösterilirse, aynı şekilde örnek büyüklüklerinin toplamı, toplam örnek büyüklüğüne eşittir. Yani,

$$n = \sum_{i=1}^k n_h \quad (5.2)$$

şeklindedir.

Zümrelere Göre Örnekleme yönteminin, tahminlerinin güvenilirliğini arttırdığı daha önce belirtilmişti. Bunun da homojen birimlerin biraraya toplanarak yapıldığı söylenmişti. Ancak bazen homojen birimleri biraraya toplamak mümkün olmayabilir. Mesela, bazen anakütlenin çeşitli bölümleri için ayrı ayrı bilgiye ihtiyaç duyulabilir. Bu durumda, yine anakütleyi zümrelere ayırıp herbir zümreyi anakütle gibi işleme tabi tutarak daha sonra esas anakütle için tahminler yapmak gerekir. Zümreleme işleminin bu avantajı çok geniş bir uygulama alanı bulmasına neden olur¹⁵. Örneğin, bir fabrikada çalışanların toplam sayısının yanında aynı sayının kadın ve erkek çalışanlar için ayrı ayrı incelenmesi istenebilir. Bu aşamada anakütleyi zümrelere ayırmanın ne kadar faydalı olduğu ortaya çıkar.

Zümrelere Göre Örnekleme yöntemi, büyük alan araştırmalarında da kolaylık sağlar. Araştırmayı yapan şirketlerin alan araştırmasının yapıldığı yerlerde kendi büroları mevcut ise, her büro kendi bölgesinde yapılan araştırmalardan sorumlu tutulduğunda maliyette de düşme sağlanacaktır.

¹⁵ Moore, D., Statistics Concepts and Controversies, New York: W.H.Freeman & Company, 1991, s. 48.

Zümrelere Göre Örneklemenin başka bir yararı ise, örneklem birimlerinin hepsine aynı usulle yaklaşmanın mümkün olmadığı durumlarda ortaya çıkar. Örneğe dahil birimlerin bazılarını telefonla, ya da anket göndererek ulaşılırken, okuma yazma bilmeyen birimler için anketörün devreye girip anketi doldurması istenir. İşte bu durumda, anakütleyi zümrelere ayırarak, yani telefonla ulaşılacakları bir zümre, anket gönderilecekleri bir zümre ve okuma yazma bilmeyenleri bir zümre oluşturacak şekilde düzenleyerek maliyette düşme sağlanabilir. Daha sonra bu zümrelerin her biri birer anakütle gibi düşünülerek basit tesadüfi usul ile birer örnek çekilir ve toplam örnek büyüklüğü belirlenir. Z.G.Ö. yöntemi zümrelere ayrılmış bu alt anakütlelere dayanarak yapılacak tahminlerden en iyi sonucu elde etmeyi amaçlar. Bu yüzden de örnek büyüklükleri olan n_h ların en doğru şekilde zümreler arasında dağıtılması gerekir.

Zümrelere Göre Örneklemde anakütle ortalaması ve tahminler bölünmesi varyansı iki ana başlık altında incelenecektir. Bunlar Momentler Yöntemi İle Tahmin ve Regresyon Yöntemi İle Tahmin'dir. Zümrelere Göre Örneklemde de anakütle parametrelerinin tahmininde daha önce açıklanan dört yöntem kullanılmaktadır. Bu yöntemlerden Maksimum Likelihood Yöntemi anakütle dağılımının bilinmesini, Nisbet Usulü ile tahmin yöntemi ise noktalar arasından geçirilecek doğrunun ya orijinden ya da orijine çok yakın bir yerden geçirilmesini gerektirmektedir. Diğer bir yöntem olan Momentler Yöntemi ile tahmin ise ancak tek bir değişken ele alındığı zaman iyi bir tahmin yöntemi olmaktadır. Fakat istatistiksel araştırmalarda bir değişken yerine birden fazla değişken ile çalışmak daha güvenilir ve daha iyi tahminler vermektedir. Bu da ancak Regresyon Yöntemi ile tahmin yaparak gerçekleştirilebilir. Bu yüzden çalışmada öncelikle Momentler Yöntemi ile tahminden bahsedilecek daha sonra da ayrıntılı olarak Regresyon Yöntemi açıklanacaktır.

5.1. Momentler Yöntemi İle Tahmin

Zümrelere Göre Örneklemede ilk olarak ele alınacak Momentler Yöntemi İle Tahmin, anakütle ortalaması ve anakütle nisbetinin tahmini ile bu tahminlerin varyansını içeren alt başlıklar şeklinde anlatılacaktır.

5.1.1. Anakütle Ortalamasının Tahmini

Tahminlerde kullanılacak notasyonu vermek açıklamaları kolaylaştıracaktır.

N_h , h. zümredeki toplam birim

n_h , h. zümreden alınan örnekteki toplam birim

$W_h = \frac{N_h}{N}$, zümre tartısı

$\bar{X}_h = \frac{\sum_{i=1}^{N_h} X_h}{N_h}$, h. zümrenin ortalaması

$\bar{x}_h = \frac{\sum_{i=1}^{n_h} x_h}{n_h}$, h. zümreden alınan örneğin ortalaması

$S_h^2 = \frac{\sum_{i=1}^{N_h} (X_h - \bar{X}_h)^2}{N_h - 1}$, h. zümrenin varyansı

Anakütle zümrelere ayrıldıktan sonra her bir zümre için bir ortalama tahmin edilir ki bu $\bar{X}_h$ ile gösterilir;

$$E(\bar{x}_h) = \bar{X}_h, \quad \bar{x}_h = \hat{\bar{X}}_h \quad (5.3)$$

Anakütle ortalaması, tahmin edilen zümre ortalamalarının tartılı ortalamasıdır ve aşağıdaki şekilde ifade edilir:

$$\bar{X} = \frac{1}{N} \sum_{h=1}^k N_h \bar{X}_h \quad (5.4)$$

Bu ifadeyi $W_h = \frac{N_h}{N}$ 'yi yerine yazmak suretiyle

$$\bar{X} = \sum W_h \bar{X}_h \quad (5.5)$$

şeklinde de elde etmek mümkündür. (5.5) nolu ifadede zümre ortalaması yerine onun tahmini değeri olan zümreden çekilen örneğin ortalaması konularak

$$\bar{x}_h = \hat{X}_h \quad (5.6)$$

özelliğinden faydalanarak anakütle ortalamasının tahmini

$$\hat{\bar{X}} = \sum W_h \bar{x}_h \quad (5.7)$$

şeklinde ifade edilebilir.

Daha önce de belirtildiği gibi Basit Tesadüfi Örneklemede anakütle dağılımı çok büyük olduğu zaman örnekleme planını uygulamak çok zaman alıcı olur. Bu yüzden, amacı anakütle varyansını küçültmek olan Zümrelere Göre Örnekleme yöntemi kullanılır. Zümrelere Göre Örneklemede tahminler bölünmesinin varyansı, örnek büyüklüğünün zümreler arasındaki paylaşımına göre üç şekilde ortaya çıkar. Bunlar sırası ile Basit Usul, Nisbet Usulü ve Optimum Usul'e göre varyanslardır.

Örnek ortalamaları teorik bölünmesinin varyansı;

$$V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{1}{N^2} \sum N_h^2 V(\bar{x}_h) \quad (5.8)$$

ile hesaplanır, buradan zümre varyansı

$$V(\bar{x}_h) = \frac{S_h^2}{n_h} \frac{N_h - n_h}{N_h} \quad (5.9)$$

yerine konduğunda; tahminler bölünmesinin varyansı

$$V(\bar{x}) = \frac{1}{N^2} \sum \frac{N_h^2 S_h^2}{n_h} \frac{N_h - n_h}{N_h} \quad (5.10)$$

$$V(\bar{x}) = \frac{1}{N^2} \sum N_h^2 \frac{S_h^2}{n_h} \quad (5.12)$$

Bu eşitlikte $W_h = \frac{N_h}{N}$ yazmak suretiyle

$$V(\bar{x}) = \sum \left(\frac{N_h}{N} \right)^2 \frac{S_h^2}{n_h} \quad (5.13)$$

$$V(\bar{x}) = \sum W_h^2 \frac{S_h^2}{n_h} \quad (5.14)$$

şeklinde ifade edilebilir.

S_h^2 bilinmediği taktirde yerine $E(s_h^2) = S_h^2$ özelliğinden faydalanarak örnek değeri konulur ve tahminler bölünmesi varyansının tahmini aşağıdaki gibi elde edilir:

$$\hat{\sigma}_{\bar{x}}^2 = \sum W_h^2 \frac{s_h^2}{n_h} \quad (5.15)$$

Bu yöntem Z.G.Ö. de Basit usul olarak adlandırılır ve örnek büyüklüğünün zümreler arasında dağıtılmasına ilişkin hiçbir kistas belirtilmeden uygulanır. Daha önce de ifade edildiği gibi Z.G.Ö. , örnek büyüklüğünün zümreler arasında en iyi şekilde dağıtılmasıyla ilgilenir. Bundan sonra yukarıda ifade edilen varyans $V_{bas}(\bar{x})$ ile gösterilecektir.

Anakütleden örneğe alınan birimlerin zümrelere dağıtılmasında kullanılan bir diğer usul de nisbet usulüdür. Adından da anlaşılacağı gibi nisbet usulünde, herbir zümrenin büyüklüğü anakütle büyüklüğüne oranlanarak zümre nisbetleri bulunur. Örnek büyüklükleri;

$$\frac{N_1}{N} = \frac{n_1}{n}; \frac{N_2}{N} = \frac{n_2}{n}; \dots; \frac{N_h}{N} = \frac{n_h}{n} \quad (5.16)$$

oranlarına göre hesaplandığında yani zümrenin anakütle içindeki oranı o zümre içinden çekilen örnek için de geçerli olduğunda ki bu aşağıdaki gibi ifade edilebilir;

$$W_h = \frac{N_h}{N} = \frac{n_h}{n} \text{ veya } W_h = \frac{n}{N} = \frac{n_h}{N_h} \quad (5.17)$$

zümre ortalamaları teorik bölünmesinin varyansı (nis) indisi ile gösterilen

$$V_{nis}(\bar{x}) = \frac{N-n}{N} \frac{1}{n} \sum \frac{N_h}{N} S_h^2 \quad (5.18)$$

veya

$$V_{nis}(\bar{x}) = \frac{N-n}{N} \frac{1}{n} \sum W_h S_h^2, \quad \frac{n}{N} > 0.05 \quad (5.19)$$

ifadesine eşit olur. Bu eşitlik $\frac{n}{N} < 0.05$ olduğunda ise

$$V_{nis}(\bar{x}) = \frac{1}{n} \sum W_h S_h^2 \quad (5.20)$$

şeklinde ifade edilir. S_h^2 bilinmediği takdirde yerine örnek s_h^2 si konulur ve varyansın tahmini değeri

$$\hat{\sigma}_{nis}(\bar{x}) = \frac{1}{n} \sum W_h s_h^2 \quad (5.21)$$

şeklinde ifade edilir.

Zümrelere Göre Örneklemde amaç, zümre içi varyansları küçük tutarak anakütle varyansını azaltmaktır. Bu yüzden anakütle değerleri birbirine yakın olan birimler biraraya toplanarak küçük varyanslı anakütlelerden yola çıkılır. Tabi bu noktada örneklemenin maliyeti de gözönünde tutulacaktır. Optimum usul ise minimum varyansı veren dağıtım usulüdür. Optimum usulde örnek büyüklüğü zümreler arasında zümre varyansları gözönünde tutularak dağıtılır ve aşağıdaki gibi ifade edilir:

$$n_h = n \frac{N_h S_h}{\sum N_h S_h} = n \frac{W_h S_h}{\sum W_h S_h} \quad (5.22)$$

ile verilen n_h değeri Z.G.Ö. de basit usule göre hesaplanan teorik bölünmenin varyansını veren formülde yerine konulduğunda (opt) indisi ile aşağıdaki gibi hesaplanır

$$V_{opt}(\bar{x}) = \frac{\sum (W_h S_h)^2}{n} \quad , \quad \frac{n}{N} < 0.05 \quad (5.23)$$

S_h bilinmediği takdirde yerine tahmini değeri konulur ve varyansın tahmini

$$\hat{\sigma}_{opt}^2(\bar{x}) = \frac{1}{n} \sum (W_h s_h)^2 \quad (5.24)$$

şeklinde elde edilir.

(5.22) no'lu formülden çıkarılabilecek sonuçlar aşağıdaki gibi ifade edilebilir:

- Zümre büyüklüğü arttıkça, o zümreden çekilen örneğin büyüklüğü de artar,
- Zümre içi değişkenlik arttıkça, o zümreden çekilen örneğin büyüklüğü de artar¹⁶.

Anakütle ortalamasına ilişkin güven aralığı ise aşağıdaki gibi bulunur:

$$\text{Anakütle Ortalaması: } \hat{\bar{X}} \pm t\sigma_{\bar{x}}$$

Burada $\hat{\bar{X}} = \sum W_h \bar{x}_h$ ile hesaplanır ve $\bar{x}_h$ yerine tahmini değeri olan $\bar{x}_h$ kullanılır.

Aralıklara ilişkin bu formülde zümre dağılımlarının normal olduğu ve $\sigma_{\bar{x}}$ nın çok iyi tanımlandığı varsayılır. $\sigma_{\bar{x}}$ nın çok iyi tanımlanması, n_h ların yeterli büyüklükte belirlenmiş olması anlamına gelir. Güven aralıkları

¹⁶ Sukhatme, P.V., a.g.e., s.109.

formüllerinde yer alan t çarpanı ise normal dağılım tablolarından belirlenmiş bir α hata payına göre belirlenir ve aralık tahmini formüllerinde yer alır¹⁷.

5.1.2. Anakütle Nisbetinin Tahmini

Zümrelere Göre Örneklemede anakütle nisbetinin tahmini için yapılacak işlemler, anakütle ortalamasının tahmininde yapılanlara oldukça benzerlik gösterir.

Zümrelere göre örnekleme yönteminde, zümre nisbetlerinin tartılı aritmetik ortalaması anakütle nisbetinin yansız bir tahminidir ve $E(p) = P$ şeklinde ifade edilir. Anakütle nisbeti;

$$P = \frac{\sum_{h=1}^k N_h P_h}{\sum_{h=1}^k N_h} \quad (5.25)$$

veya

$$P = \sum W_h P_h, \quad \sum W_h = 1 \quad (5.26)$$

ile ifade edilir. Zümre nisbeti bilinmediği takdirde yerine onun tahmini değeri olan örnek nisbeti olan p_h konulur ve anakütle nisbetinin tahmini,

$$\hat{P} = \frac{\sum N_h p_h}{\sum N_h} = \sum W_h p_h \quad (5.27)$$

elde edilir.

Anakütle ortalamasının tahmininde bahsedildiği gibi, anakütle nisbetinin tahmininde de örnek büyüklüğü zümreler arasında üç şekilde dağıtılır ve varyans bu üç usule göre belirlenir.

¹⁷ H., M., M., W.G. and H., W.N., Sample Survey, New York: John Wiley & Sons, 1993, s.233.

a. Basit Usul

Örnek büyüklüğü zümreler arasında arzuya göre yani hiçbir kısıta bağlanmadan dağıtılır. Zümre nisbetlerinden oluşan teorik bölünmenin varyansı basit usule göre $P_h + Q_h = 1$ olmak üzere

$$V_{\text{bas}}(p) = \frac{1}{N^2} \sum N_h (N_h - n_h) \frac{P_h Q_h}{n_h} \quad , \quad \frac{n}{N} > 0.05 \quad (5.28)$$

şeklinde hesaplanır. Burada $W_h = \frac{N_h}{N}$ I yerine koymak ve $\frac{n}{N} < 0.05$ olmak üzere formül;

$$V_{\text{bas}}(p) = \sum W_h^2 \frac{P_h Q_h}{n_h} \quad (5.29)$$

şeklinde de ifade edilebilir.

P_h ve Q_h değerleri bilinmediğinde yerine tahmini değerleri olan p_h ve q_h konulur ve zümre nisbetleri teorik bölünmesi varyansının tahmini değeri;

$$\hat{\sigma}_{\text{bas}}^2(p) = \sum W_h^2 \frac{p_h q_h}{n_h} \quad (5.30)$$

şeklinde ifade edilir.

b. Nisbet Usulü

Örnek zümreler arasında zümre tartılarına göre dağıtılır ve varyans

$$V_{\text{nis}}(p) = \frac{N-n}{N} \frac{1}{n} \sum W_h P_h Q_h \quad , \quad \frac{n}{N} > 0.05 \quad (5.31)$$

$$V_{\text{nis}}(p) = \frac{1}{n} \sum W_h P_h Q_h \quad , \quad \frac{n}{N} < 0.05 \quad (5.32)$$

şeklinde hesaplanır.

P_h ve Q_h değerleri bilinmediği takdirde yerine tahmini değerleri konularak varyansın tahmini elde edilir:

$$\hat{\sigma}_{ms}^2(p) = \frac{1}{n} \sum W_h p_h q_h \quad (5.33)$$

c. Optimum Usul

Örnek büyüklüğü zümreler arasında en küçük varyansı verecek

$$n_h = n \frac{N_h \sqrt{P_h Q_h}}{\sum N_h \sqrt{P_h Q_h}} \quad (5.34)$$

veya

$$n_h = n \frac{W_h \sqrt{P_h Q_h}}{\sum W_h \sqrt{P_h Q_h}} \quad (5.35)$$

oranlarına göre dağıtıldığında

$$V_{opt}(p) = \frac{\left(\sum W_h \sqrt{P_h Q_h} \right)^2}{n} \quad (5.36)$$

ile hesaplanır.

Anakütle değerleri olan P_h ve Q_h lar bilinmediği takdirde yerine tahmini değerleri olan p_h ve q_h lar yazılır ve varyansın tahmini değeri elde edilir:

$$V_{opt}(p) = \frac{\left(\sum W_h \sqrt{p_h q_h} \right)^2}{n} \quad (5.37)$$

5.2. Regresyon Yöntemi İle Tahmin

Zümrelere Göre Örneklemede regresyon yöntemi ile tahminin nasıl uygulandığını anlatmadan önce regresyon yöntemi ile tahmin başlığı altında incelenen bölümden bazı hatırlatmalar yapmak yararlı olur. Daha önce de ifade edildiği gibi:

Anakütle regresyon denklemi;

$$Y = \alpha + \beta X \quad (5.38)$$

Anakütleden çekilen n büyüklüğündeki örnek regresyon denklemi;

$$y = a + bx \quad (5.39)$$

ve anakütle parametresi olan ortalamanın tahmininde kullanılan regresyon tahmin denklemi

$$\hat{\bar{Y}} = \bar{y} + b(\bar{X} - \bar{x}) \quad \text{veya} \quad (5.40)$$

$b = r \frac{s_y}{s_x}$ kullanıldığında

$$\hat{\bar{Y}} = \bar{y} + r \frac{s_y}{s_x} (\bar{X} - \bar{x}) \quad (5.41)$$

şeklinde idi. Tahmin denkleminde b değeri;

$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad (5.42)$$

ile hesaplanmakta ve $E(b) = \beta$ olduğu bilinmekte idi.

Bu bölümde, Zümrelere Göre Örnekleme içinde yer alan regresyon yöntemi iki başlık altında incelenecektir. Bunlar, ayrı ve birleşik tahmin yöntemleridir.

Burada hemen ayrı ve bileşik tahminlerin ne anlama geldiğini açıklamakta fayda vardır.

Ayrı tahminden kasıt, anakütle regresyon katsayısı olan β nın zümreden zümreye değişmesi ve buna bağlı olarak b_h değerlerinin de değişmesidir. Böylece elimizde kaç zümre var ise, o kadar regresyon denklemi olacak ve buradan anakütle parametresi tahminine geçilecektir.

Bileşik tahminde ise; anakütle regresyon katsayısı olan β , tüm zümreler için aynıdır ve buna bağlı olarak zümreden elde edilen b_h değerlerinin de aynı olduğu yani

$$b_1 = b_2 = \dots = \beta \quad (5.43)$$

olduğu kabul edilir. Böylece tek bir regresyon denklemi yardımıyla anakütle parametresi tahmin edilir.

5.2.1. Anakütle Ortalamasının Ayrı Tahmini

Zümrelere Göre Örneklemede, herbir zümreden yapılan doğrusal regresyon tahmini, zümre ağırlıkları ile çarpılıp toplanarak ayrı regresyon tahmini elde edilir. h. zümrenin regresyon tahmini ;

$$\bar{y}_{drh} = \bar{y}_h + b_h (\bar{X}_h - \bar{x}_h) \quad (5.44)$$

şeklinde ifade edilir, $\bar{y}_{drh}$ ise herbir zümrenin doğrusal regresyon denklemini gösterir. b_h katsayısı , herbir zümre için $\bar{x}_h$ da meydana gelen bir birimlik artışın $\bar{y}_{drh}$ da yarattığı ortalama değişimin göstergesidir ve zümre parametresi olan β_h ın en küçük kareler tahminidir. Ayrı regresyon tahmininden kasıt ise β_h ın zümreden zümreye değişmesidir. Anakütle parametresi olan β_h lar değiştiği için b_h değerleri de değiştiğinden dolayı tahmine ayrı regresyon

tahmini denilir. Yukarıda ifade edilen $\bar{y}_{drh}$ tahmin denkleminde, $b_h = r_h \frac{s_{yh}}{s_{xh}}$ yerine yazılırsa ayrı regresyon tahmini;

$$\bar{y}_{drh} = \bar{y}_h + r_h \frac{s_{yh}}{s_{xh}} (\bar{X}_h - \bar{x}_h) \quad (5.45)$$

şekline dönüşür.

Anakütle parametresi olan ortalamanın ayrı tahmini;

$$\bar{Y}_{dra} = \sum_{h=1}^k W_h \bar{y}_{drh} = \frac{\sum_{h=1}^k N_h \bar{y}_{drh}}{N} \quad (5.46)$$

ve anakütle toplamının ayrı regresyon tahmini

$$\hat{Y}_{dra} = N \bar{y}_{dra} \quad (5.47)$$

şeklinde bulunur.

$\bar{y}_{drh}$, $\bar{Y}_h$ in yansız bir tahmini ve $\bar{y}_{dra}$ da $\bar{Y}$ nin yansız bir tahminidir.

5.2.2. Anakütle Ortalaması Ayrı Tahmininin Varyansı

Tahmin, zümrelerde birbirinden bağımsız yapıldığı takdirde anakütle ortalaması ayrı tahmininin varyansı;

$$\begin{aligned} V(\bar{y}_{dra}) &= \sum W_h^2 V(\bar{y}_{drh}) \\ &= \sum \frac{W_h^2}{n_h} (1 - f_h) (S_{yh}^2 - 2b_h S_{yxh} + b_h^2 S_{xh}^2) \end{aligned} \quad (5.48)$$

şeklinde ifade edilir.

$$b_h = \beta_h = \frac{S_{yxh}}{S_{xh}^2} \quad (5.49)$$

olduğunda varyans en küçük değere sahip olur ve varyansın minimum değerini bulmak için bu ifade (5.48) da yerine konulur ve böylece minimum varyans

$$V_{\min}(\bar{y}_{dra}) = \sum \frac{W_h^2}{n_h} (1 - f_h) S_{yh}^2 (1 - r_h^2) \quad (5.50)$$

şeklini alır. Anakütle parametreleri olan S_{yh}^2, S_{yxh} ve S_{xh}^2 bilinmediği takdirde yerlerine örnek değerleri konur ve varyansın tahmini

$$\hat{\sigma}^2(\bar{y}_{dra}) = \sum \frac{W_h^2}{n_h} (1 - f_h) (s_{yh}^2 - 2b_h s_{yxh} + b_h^2 s_{xh}^2) \quad (5.51)$$

şeklinde elde edilir.

Anakütle ortalaması ayrı tahmininin varyans tahmininde yer alan ve s_{yxh} , s_{xh}^2 ve s_{yh}^2 değerlerinin hesabı ise aşağıdaki gibidir:

$$s_{yxh} = \frac{\sum (x_h - \bar{x}_h)(y_h - \bar{y}_h)}{n_h - 1} \quad (5.52)$$

$$s_{xh}^2 = \frac{\sum (x_h - \bar{x}_h)^2}{n_h - 1} \quad (5.53)$$

$$s_{yh}^2 = \frac{\sum (y_h - \bar{y}_h)^2}{n_h - 1} \quad (5.54)$$

5.2.3. Anakütle Ortalamasının Bileşik Tahmini

Anakütle regresyon parametresi olan β_h ların birbirine yakın olduğu durumda ise örnekten elde edilen b_h larında anakütlerde aynı olduğu kabul edilebilir. Bu durumda, bileşik regresyon tahmini¹⁸;

$$\hat{\bar{Y}}_{drb} = \bar{y}_b + b(\bar{X} - \bar{x}_b) \quad (5.55)$$

ile yapılır. Denklemden yeralan

$$\bar{y}_b = \sum W_h \bar{y}_h : \text{z.g.ö. de ortalama tahmini}$$

$$\bar{x}_b = \sum W_h \bar{x}_h : \text{z.g.ö. de ortalama tahmini}$$

dir.

Her zümrede b değerleri aynı olduğunda $\bar{y}_{drb}$, $\bar{Y}$ nin yansız bir tahminidir.

¹⁸ Çingı, H., a.g.e., s.187.

5.2.4. Anakütle Ortalaması Bileşik Tahmininin Varyansı

Ortalama tahminlerine ilişkin varyans ise;

$$V(\bar{y}_{drb}) = \sum \frac{W_h^2}{n_h} (1 - f_h) (S_{yh}^2 - 2bS_{yxh} + b^2 S_{xh}^2) \quad (5.56)$$

şeklinde hesaplanır.

Anakütle parametreleri olan S_{yh}^2, S_{yxh} ve S_{xh}^2 bilinmediği takdirde yerlerine örnek değerleri konur ve varyansın tahmini

$$\hat{\sigma}^2(\bar{y}_{drb}) = \sum \frac{W_h^2}{n_h} (1 - f_h) (s_{yh}^2 - 2bs_{yxh} + b^2 s_{xh}^2) \quad (5.57)$$

şeklinde bulunur.

β_h , zümre regresyon katsayısı tüm zümrelerde aynı ise, ayrı regresyon tahminine ilişkin varyans bileşik tahmine ilişkin varyanstan daha küçüktür .

5.3. Zümrelere Göre Örneklemede Regresyon İle Tahmin

Yöntemlerinin Karşılaştırılması

Bu karşılaştırmaya girmeden önce, örnek zümre içinden Zümrelere Göre Örnekleme Yöntemine göre değilde Basit Tesadüfi Örnekleme Yöntemine göre çekilseydi varyansın tahmini değeri nasıl bulunacaktı onu görelim.

Hatırlanacak olursa B.T.Ö. yöntemine göre bulunan örnek ortalamaları teorik bölünmesinin varyansı N nin büyük değerleri için

$$V(\bar{x}) = \frac{\sigma^2}{n} \quad (5.58)$$

ve $S^2 = \sigma^2 \frac{N}{N-1}$ idi. N nin büyüyen değerleri için $S^2 = \sigma^2$ yazılabildiğinden dolayı

$$V(\bar{x}) = \frac{S^2}{n} \quad (5.59)$$

şeklinde de ifade ediliyordu.

Örnek zümrelerden, Z.G.Ö. yöntemine göre değilde B.T.Ö. yöntemine göre çekilse idi, varyansın tahmini değeri

$$\hat{S}^2 = \hat{\sigma}^2 = \frac{1}{N} \sum N_h^2 S_h^2 + \frac{1}{N} \sum N_h (\bar{x}_h - \bar{X})^2 \quad (5.60)$$

veya

$$\hat{\sigma}^2 = \sum W_h S_h^2 + \sum W_h (\bar{x}_h - \bar{X})^2 \quad (5.61)$$

şeklinde olur. Daha sonra bulunan bu değer (5.58) no'lu denklemde yerine konularak B.T.Ö. yöntemine göre çekilmiş örneğin varyansı elde edilir. (5.60)' nolu de yeralan ilk terim; zümre varyanslarının tartılı aritmetik ortalaması, ikinci terim ise zümre ortalamalarının varyansını ifade eder.

Zümrelere Göre Örnekleme yönteminden elde edilen varyans ile Basit Tesadüfi Örnekleme yönteminden elde edilen varyanslar birbirlerine oranlanarak etkinlik karşılaştırması yapılarak sonuca varılır. Etkinlik hesabı ise aşağıdaki gibi yapılır:

$$\text{Etkinlik} = \frac{V(\text{B.T.Ö.})}{V(\text{Z.G.Ö.})} - 1 \quad (5.62)$$

Bu kısa açıklamadan sonra Zümrelere Göre Örneklemede ayrı ve bileşik regresyon yöntemlerine göre yapılan tahminlerden hangisinin daha etkin olduğunu yani hangisinin varyansının daha küçük olduğunu görelim.

Ayrı regresyon ile tahminde varyans minimum değerini $b_h = \beta_h$ olduğu zaman alır. Bu ifade daha önce (5.50) nolu formül ile verilmiştir.

Varyansın minimum değeri, bileşik regresyon tahmini yöntemi kullanıldığında ise (5.56) nolu formülde b yerine

$$B_b = \frac{\sum_h W_h^2 (1-f_h) s_{yxh} / n_h}{\sum_h W_h^2 (1-f_h) s_{xh}^2 / n_h} \quad (5.63)$$

değeri konularak bulunur.

B_b değeri, zümre regresyon katsayısının ağırlıklı ortalaması olarak adlandırılır ve

$$B_h = \frac{S_{yxh}}{S_{xh}^2} \quad (5.64)$$

ile ifade edilirse ve (5.63) nolu formüldeki

$$\frac{W_h^2 (1-f_h)}{n_h} s_{xh}^2 \quad (5.65)$$

ifadesine a_h denilirse;

$$B_b = \frac{\sum_h W_h^2 (1-f_h) s_{yxh} / n_h}{\sum_h W_h^2 (1-f_h) s_{xh}^2 / n_h} = \frac{\sum_h a_h B_h}{\sum_h a_h} \quad (5.66)$$

olur.

B_b değeri, ayrı ve bileşik regresyon ile tahmin yöntemlerinde ortalama tahmininin varyansını veren formüllerde yerine konulup fark alınırsa;

$$\begin{aligned} v_{\min}(\bar{y}_{drb}) - v_{\min}(\bar{y}_{dra}) &= -2 \frac{(\sum a_h b_h)^2}{\sum a_h} + (\sum a_h) B_h^2 + \sum a_h B_h^2 \\ &= -2 \frac{(\sum a_h b_h)^2}{\sum a_h} + (\sum a_h) B_h^2 + \sum a_h B_h^2 \\ &= \sum a_h B_h^2 - 2B_b \sum a_h B_h + \sum a_h B_b^2 \\ &= \sum a_h (B_h - B_b)^2 \end{aligned} \quad (5.67)$$

elde edilir.

$B_h \neq B_b$ yani tüm zümreler için B_h zümre regresyon katsayısı aynı olduğu zaman ayrı tahmine göre bulunan varyans, biileşik tahmine göre bulunan varyanstan küçük çıkar ve

$$v_{\min}(\bar{y}_{dra}) < v_{\min}(\bar{y}_{drb}) \quad (5.68)$$

şeklinde ifade edilir.



6. UYGULAMA

Çok uluslu bir şirket olan (A) da çalışanlar aldıkları haftalık ücretlere göre, birbirine yakın ücret alanların biraraya getirildiği 4 grup şeklinde toplanmıştır. Elde edilen bu grupların en önemli özelliği kendi içlerinde homojen, kendi aralarında heterojen olmalarıdır. Diğer bir ifade ile, her grup zümre şeklinde adlandırılmakta, elde edilen zümrelerin araştırmaya esas teşkil edecek vasıf açısından yekdiğerinden farklı olduğu düşünülmektedir. Kısaca, istatistik tanım olarak, zümre ortalamaları birbirinden farklıdır.

Oluşturulan bu zümrelerin zümre tartıları olan W_h 'lar sırasıyla 0.10; 0.50; 0.25 ve 0.15'dir. Bu zümreler Basit Tesadüfi Örneklemeye yöntemiyle örneklemeye tabi tutulmuştur. Yani her grup bağımsız bir anakütle gibi düşünülmüş, bu gruba dahil her birime, çekilecek örneğe girme bakımından eşit şans tanınmaya çalışılmıştır. Bunun için bilgisayar tarafından oluşturulan tesadüfi sayılar tablosu içinden örnek birimler çekilmiş ve sonuçta toplam anakütle için tesbit edilen 400 örnek işçinin zümre tartılarına göre zümre dağılımına uygun 40; 200; 100 ve 60 birimli alt örnekler oluşturulmuştur.

Böylece tesadüfi sayılara göre oluşturulan listelerden belirlenen örnek işçilerin ad, soyad ve çalıştıkları bölümler, hazırlanan soru kağıtlarının üzerine yazılarak adı geçen işçilere dağıtılmıştır. Toplanan verilere göre hesaplanan örnek değerleri aşağıdaki Tablo:1'de verilmiştir.

Tablo:1

Örnek Değerleri

N_h	n_h	$\bar{y}_h$	s_h^2	W_h
10.000	40	8	90.0	0.10
50.000	200	16	19.6	0.50
25.000	100	24	62.5	0.25
15.000	60	30	32.4	0.15

1. Bilinmeyen parametre olan anakütle ortalamasının tahmini değeri aşağıdaki gibi hesaplanır:

$$\hat{\bar{Y}} = \sum W_h \bar{y}_h$$

$$\hat{\bar{Y}} = [(0.10 \times 8) + (0.50 \times 16) + (0.25 \times 24) + (0.15 \times 30)]$$

$$= 19.3 \text{ milyonTL / hafta}$$

2. Tahminler bölünmesinin varyansı veya diğer bir ifade ile anakütle parametresi tahmininin standart hatasının karesi:

$$V(\bar{y}) = \hat{\sigma}_{\bar{y}}^2 = \frac{1}{n} \sum W_h s_h^2$$

$$= \frac{1}{400} [(0.10 \times 90.0) + (0.50 \times 19.6) + (0.25 \times 62.5) + (0.15 \times 32.4)]$$

$$= \frac{39.285}{400} = 0.0982$$

$$\hat{\sigma}_{\bar{y}} = 0.313 \text{ milyonTL / hafta}$$

3. Bilinmeyen anakütle parametresi olan $\bar{Y}$ 'yı içeren aralıklardan biri olma olasılığı 0.95 olan aralık:

$$\hat{\bar{Y}}_{1,2} = \hat{\bar{Y}} \pm t\sigma_{\bar{y}}$$

$$\hat{\bar{Y}}_{1,2} = 19.3 \pm (1.96)(0.313)$$

$$\hat{\bar{Y}}_1 = 19.914 \text{ milyonTL / hafta}$$

$$\hat{\bar{Y}}_2 = 18.686 \text{ milyonTL / hafta}$$

4. 400 birimli bu örnek, incelenen anakütle zümrelere bölünmeksizin, tüm işçilerin oluşturduğu anakütleden, yine üretilen tesadüfî sayılar tablosunu kullanarak çekilmiş olsa idi bu durumda tahminler bölünmesinin varyansı ne olacaktı ve sonuçta zümreleri oluşturmak suretiyle katlandığımız maliyetin bize yararı ne olmuştur, onu bulmaya çalışalım:

Bunun için önce anakütle varyansını tahmin edelim, sonra bu tahmini varyansı kullanarak B.T.Ö. yöntemine dayalı tahminler bölünmesinin varyansını bulalım ve son olarak da etkinliği hesaplayalım.

$$\hat{S}^2 = \hat{\sigma}^2 = \frac{1}{N} \sum N_h s_h^2 + \frac{1}{N} \sum N_h (\bar{y}_h - \bar{Y})^2 \quad \text{veya}$$

$$= \sum W_h s_h^2 + \sum W_h (\bar{y}_h - \bar{Y})^2$$

$$\sum W_h s_h^2 = [(0.10)(90.0) + (0.50)(19.6) + (0.25)(62.5) + (0.15)(32.4)]$$

$$= 39.285$$

$$\sum W_h (\bar{y}_h - \bar{Y})^2 = [(0.10)(8 - 19.3)^2 + (0.50)(16 - 19.3)^2 + (0.25)(24 - 19.3)^2 + (0.15)(30 - 19.3)^2]$$

$$= 40.91$$

$$\hat{\sigma}^2 = 39.285 + 40.91 = 80.195$$

$$\hat{\sigma}_y^2 = \frac{80.195}{400} = 0.20 \text{ milyon TL/hafta}$$

Z.G.Ö. yönteminin B.T.Ö. yöntemine etkinliği ise aşağıdaki gibi bulunur:

$$\text{Etkinlik} = \frac{V(\text{B.T.Ö.})}{V(\text{Z.G.Ö.})} - 1 = \frac{0.2000}{0.0982} - 1 = 1.0367$$

5. Şimdi B.T.Ö. yöntemine göre bulunan tahminler bölünmesi varyansını kullanarak anakütle parametresi olan $\bar{Y}$ yı içeren aralıklardan biri olma olasılığı 0.95 olan bir aralık bulalım:

B.T.Ö. yöntemine göre bulunan tahminler bölünmesi varyansı 0.20 idi. Bu değeri aralık tahmini formülünde yerine koyarsak:

$$\hat{\bar{Y}}_{1,2} = \hat{\bar{Y}} \pm t\sigma_{\bar{y}}$$

$$\hat{\bar{Y}}_{1,2} = 19.3 \pm (1.96)(0.45)$$

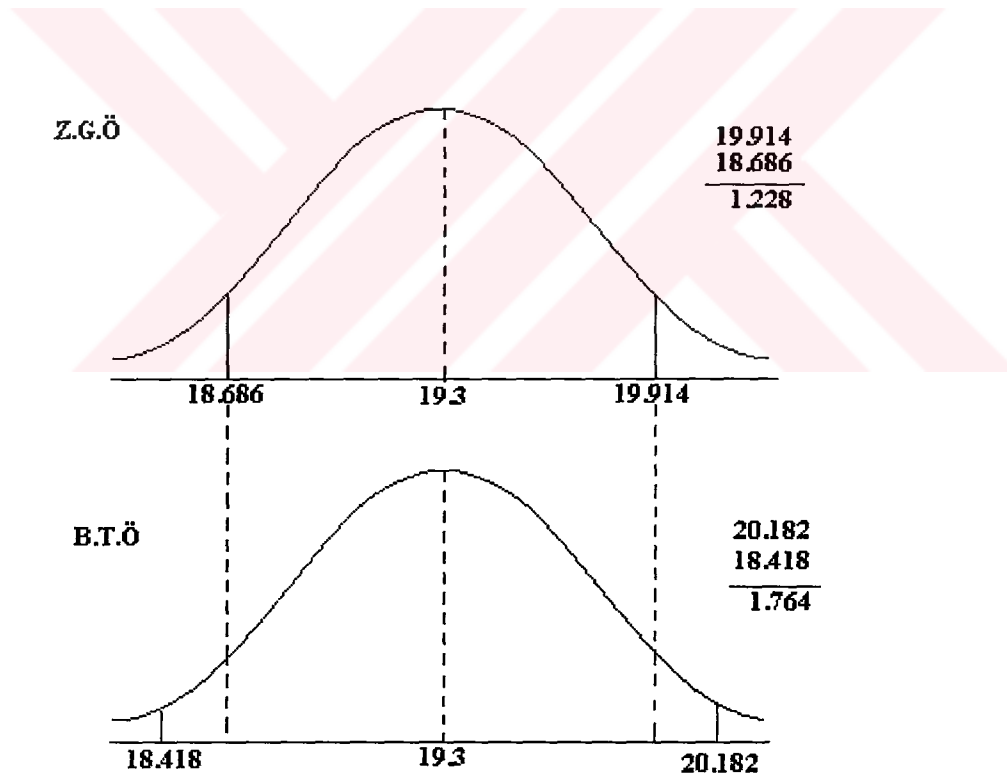
$$\hat{Y}_1 = 20.182 \text{ milyonTL / hafta}$$

$$\hat{Y}_2 = 18.418 \text{ milyonTL / hafta}$$

Z.G.Ö. yöntemine göre bulunan aralıklar sırasıyla $\hat{Y}_1 = 19.914$ milyon TL/hafta ve $\hat{Y}_2 = 18.686$ milyon TL/hafta idi. Her iki yöntem için elde edilen bu aralıkları Şekil:1 üzerinde göstermek yöntemleri karşılaştırmak açısından faydalı olur.

Şekil 1:

**Z.G.Ö. ile B.T.Ö. Yöntemlerinin
Karşılaştırılması**



Görüldüğü gibi, Z.G.Ö. yöntemini kullanmak, B.T.Ö. yönteminden daha faydalıdır. Çünkü, anakütle parametresine ilişkin aralık Z.G.Ö. yöntemine göre bulunduğu daha dar olduğu görülür.

Örnekleme yönteminde etkinlik, tahminler bölünmesi varyansının ne kadar küçültülebildiğine bağlıdır. Yukarıda bu özellik örnekleme yöntemini değiştirmek suretiyle yapılmıştır.

6. Şimdi sırada etkinliği arttırmanın bir başka yolu olan tahmin yöntemini değiştirmeye gelmiştir. Bu iş momentler yöntemi yerine regresyon yöntemini kullanmak suretiyle gerçekleştirilmiştir.

Regresyon yöntemi ile tahmin için gerekli olan veriler her zümre için ayrı bir tablo şeklinde toplanmış ve hesaplar her zümre için ayrı regresyon tahmini yöntemiyle yapılmıştır. Sıra ile 4 zümre için veriler ve gerekli hesaplar aşağıda sunulmuştur:



Tablo:2

Zümre 1

n = 40	$\sum X = 140$	$\sum Y = 320$	$\sum XY = 2170.22$
	$\sum X^2 = 1769.2$	$\sum Y^2 = 6070$	

$$\bar{X} = \frac{\sum X}{n} = \frac{140}{40} = 3.5$$

$$V(x) = \sigma_x^2 = \frac{\sum X^2}{n} - \bar{X}^2$$
$$\sigma_x^2 = \frac{1769.2}{40} - (3.5)^2 = 31.98$$

$$s_{x1}^2 = \frac{\sigma^2 n}{n-1} = \frac{31.98 \times 40}{39} = 32.8$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{320}{40} = 8$$

$$V(y) = \sigma_y^2 = \frac{\sum Y^2}{n} - \bar{Y}^2$$
$$\sigma_y^2 = \frac{6070}{40} - (8)^2 = 87.75$$

$$s_{y1}^2 = \frac{\sigma^2 n}{n-1} = \frac{87.75 \times 40}{39} = 90.0$$

$$b_{x1} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2} = \frac{2170.22 - 40(3.5)(8)}{1769.2 - 40(12.25)} = 0.821$$

$$b_{y1} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum Y^2 - n\bar{Y}^2} = \frac{2170.22 - 40(3.5)(8)}{6070 - 40(64)} = 0.30$$

$$r_1 = \sqrt{b_{x1} b_{y1}} = \sqrt{0.821 \times 0.30} = 0.496$$

Zümre 1 için $\bar{X}_1 = 5$ olarak verilmiştir. Buradan yola çıkarak Zümre 1 için ortalamanın doğrusal regresyon ile tahmini aşağıdaki gibidir:

$$\hat{\bar{Y}}_{dr1} = \bar{y}_1 + b_1 (\bar{X}_1 - \bar{x}_1)$$

$$= 8 + 0.821(5 - 3.5)$$

$$= 9.232$$

$$b_1 = r_1 \frac{s_{y_1}}{s_{x_1}}$$

ifadesi denklemde yerine konulursa Zümre 1 için doğrusal regresyon tahmini aşağıdaki gibi bulunur:

$$\hat{\bar{Y}}_{dr1} = \bar{y}_1 + r_1 \frac{s_{y_1}}{s_{x_1}} (\bar{X}_1 - \bar{x}_1)$$

$$= 8 + 0.496 \frac{9.487}{5.727} (5 - 3.5)$$

$$= 9.232$$

Tablo:3

Zümre 2

n = 200	$\sum X = 2560$	$\sum Y = 3200$	$\sum XY = 42788.8$
	$\sum X^2 = 36010$	$\sum Y^2 = 55100$	

$$\bar{X} = \frac{\sum X}{n} = \frac{2560}{200} = 12.8$$

$$V(x) = \sigma_x^2 = \frac{\sum X^2}{n} - \bar{X}^2$$
$$\sigma_x^2 = \frac{36010}{200} - (12.8)^2 = 16.21$$

$$s_{x_2}^2 = \frac{\sigma^2 n}{n-1} = \frac{16.21 \times 200}{199} = 16.29$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{3200}{200} = 16$$

$$V(y) = \sigma_y^2 = \frac{\sum Y^2}{n} - \bar{Y}^2$$
$$\sigma_y^2 = \frac{55100}{200} - (16)^2 = 19.5$$

$$s_{y_2}^2 = \frac{\sigma^2 n}{n-1} = \frac{19.5 \times 200}{199} = 19.6$$

$$b_{x_2} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2} = \frac{42788.8 - 200(12.8)(16)}{36010 - 200(12.8)^2} = 0.564$$

$$b_{y_2} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum Y^2 - n\bar{Y}^2} = \frac{42788.8 - 200(12.8)(16)}{55100 - 200(16)^2} = 0.469$$

$$r_2 = \sqrt{b_{x_2} b_{y_2}} = \sqrt{0.564 \times 0.469} = 0.514$$

Zümre 2 için $\bar{X}_2 = 15$ olarak verilmiştir ve zümre 2 için doğrusal regresyon ile tahmini aşağıdaki gibidir:

$$\begin{aligned}\hat{\bar{Y}}_{dr2} &= \bar{y}_2 + b_2(\bar{X}_2 - \bar{x}_2) \\ &= 16 + 0.564(15 - 12.8) \\ &= 17.241\end{aligned}$$

$$b_2 = r_2 \frac{s_{y_2}}{s_{x_2}}$$

ifadesi denklemde yerine konulursa Zümre 1 için doğrusal regresyon tahmini aşağıdaki gibi bulunur:

$$\begin{aligned}\hat{\bar{Y}}_{dr2} &= \bar{y}_2 + r_2 \frac{s_{y_2}}{s_{x_2}}(\bar{X}_2 - \bar{x}_2) \\ &= 16 + 0.514 \frac{4.427}{4.036}(15 - 12.8) \\ &= 17.241\end{aligned}$$

Tablo:4

Zümre 3

n = 100	$\sum X = 1020$	$\sum Y = 2400$	$\sum XY = 26500.5$
	$\sum X^2 = 12586$	$\sum Y^2 = 63788$	

$$\bar{X} = \frac{\sum X}{n} = \frac{1020}{100} = 10.2$$

$$V(x) = \sigma_x^2 = \frac{\sum X^2}{n} - \bar{X}^2$$
$$\sigma_x^2 = \frac{12586}{100} - (10.2)^2 = 21.82$$

$$s_{x_3}^2 = \frac{\sigma^2 n}{n-1} = \frac{21.82 \times 100}{99} = 22.04$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{2400}{100} = 24$$

$$V(y) = \sigma_y^2 = \frac{\sum Y^2}{n} - \bar{Y}^2$$
$$\sigma_y^2 = \frac{63788}{100} - (24)^2 = 61.88$$

$$s_{y_3}^2 = \frac{\sigma^2 n}{n-1} = \frac{61.88 \times 100}{99} = 62.5$$

$$b_{x_3} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2} = \frac{26500.5 - 100(10.2)(24)}{12586 - 100(10.2)^2} = 0.926$$

$$b_{y_3} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum Y^2 - n\bar{Y}^2} = \frac{26500.5 - 100(10.2)(24)}{63788 - 100(24)^2} = 0.326$$

$$r_3 = \sqrt{b_{x_3} b_{y_3}} = \sqrt{0.926 \times 0.326} = 0.55$$

Zümre 3 için $\bar{X} = 12$ olarak verilmiştir. Buradan yola çıkarak Zümre 3 için doğrusal regresyon tahmini aşağıdaki gibidir:

$$\begin{aligned}\hat{Y}_{dr3} &= \bar{y}_3 + b_3(\bar{X}_3 - \bar{x}_3) \\ &= 24 + 0.926(12 - 10.2) \\ &= 25.667\end{aligned}$$

$$b_3 = r_3 \frac{s_{y_3}}{s_{x_3}}$$

ifadesi denklemde yerine konulursa Zümre 3 için doğrusal regresyon tahmini aşağıdaki gibi bulunur:

$$\begin{aligned}\hat{Y}_{dr3} &= \bar{y}_3 + r_3 \frac{s_{y_3}}{s_{x_3}}(\bar{X}_3 - \bar{x}_3) \\ &= 24 + 0.55 \frac{7.906}{4.695}(12 - 10.2) \\ &= 25.667\end{aligned}$$

Tablo:5

Zümre 4

n = 60	$\sum X = 918$	$\sum Y = 1800$	$\sum XY = 28097.9$
	$\sum X^2 = 15322.2$	$\sum Y^2 = 55911.6$	

$$\bar{X} = \frac{\sum X}{n} = \frac{918}{60} = 15.3$$

$$V(x) = \sigma_x^2 = \frac{\sum X^2}{n} - \bar{X}^2$$
$$\sigma_x^2 = \frac{15322.2}{60} - (15.3)^2 = 21.28$$

$$s_{x4}^2 = \frac{\sigma^2 n}{n-1} = \frac{21.28 \times 60}{59} = 21.64$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{1800}{60} = 30$$

$$V(y) = \sigma_y^2 = \frac{\sum Y^2}{n} - \bar{Y}^2$$
$$\sigma_y^2 = \frac{55911.6}{60} - (30)^2 = 31.86$$

$$s_{y4}^2 = \frac{\sigma^2 n}{n-1} = \frac{31.86 \times 60}{59} = 32.4$$

$$b_{x4} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2} = \frac{28097.9 - 60(15.3)(30)}{15322.2 - 60(15.3)^2} = 0.437$$

$$b_{y4} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum Y^2 - n\bar{Y}^2} = \frac{28097.9 - 60(15.3)(30)}{55911.6 - 60(30)^2} = 0.292$$

$$r_4 = \sqrt{b_{x4} b_{y4}} = \sqrt{0.437 \times 0.292} = 0.357$$

Zümre 4 için $\bar{X}_4 = 17$ olarak verilmiştir. Buradan yola çıkarak Zümre 4 için doğrusal regresyon tahmini aşağıdaki gibidir:

$$\begin{aligned}\hat{\bar{Y}}_{dr4} &= \bar{y}_4 + b_4(\bar{X}_4 - \bar{x}_4) \\ &= 30 + 0.437(17 - 15.3) \\ &= 30.743\end{aligned}$$

$$b_4 = r_4 \frac{s_{y_4}}{s_{x_4}}$$

ifadesi denklemde yerine konulursa Zümre 4 için doğrusal regresyon tahmini aşağıdaki gibi bulunur:

$$\begin{aligned}\hat{\bar{Y}}_{dr4} &= \bar{y}_4 + r_4 \frac{s_{y_4}}{s_{x_4}} (\bar{X}_4 - \bar{x}_4) \\ &= 30 + 0.357 \frac{5.692}{4.652} (17 - 15.3) \\ &= 30.743\end{aligned}$$

4 zümre için yapılan doğrusal regresyon tahmin denklemlerini de bulduktan sonra anakütle ortalamasının tahminine geçilebilir:

$$\begin{aligned}\hat{\bar{Y}}_{dra} &= \sum_{h=1}^4 W_h \hat{\bar{Y}}_h \\ &= (0.10 \times 9.232) + (0.50 \times 17.241) + (0.25 \times 25.667) + (0.15 \times 30.743) \\ &= 20.5719\end{aligned}$$

Anakütle ortalaması ayrı tahmininin varyans tahmini;

$$\hat{\sigma}^2(\bar{y}_{dra}) = \sum_{h=1}^4 \frac{W_h^2}{n_h} (1 - f_h) (s_{yh}^2 - 2b_h s_{yhx} + b_h^2 s_{xh}^2)$$

şeklindedir. Bu ifadede yer alan s_{yhx} değerlerinin hesabı aşağıdaki gibidir:

$$s_{yx1} = \frac{\sum (x_1 - \bar{x}_1)(y_1 - \bar{y}_1)}{n_1 - 1} = \frac{1050.22}{39} = 26.93$$

$$\sum (x_1 - \bar{x}_1)(y_1 - \bar{y}_1) = \sum x_1 y_1 - n \bar{x}_1 \bar{y}_1 = 2170.22 - 40(3.5)(8) = 1050.22$$

$$s_{yx2} = \frac{\sum (x_2 - \bar{x}_2)(y_2 - \bar{y}_2)}{n_2 - 1} = \frac{1828.5}{199} = 9.19$$

$$\sum (x_2 - \bar{x}_2)(y_2 - \bar{y}_2) = \sum x_2 y_2 - n \bar{x}_2 \bar{y}_2 = 42788.8 - 200(12.8)(16) = 1828.8$$

$$s_{yx3} = \frac{\sum (x_3 - \bar{x}_3)(y_3 - \bar{y}_3)}{n_3 - 1} = \frac{2020.5}{99} = 20.41$$

$$\sum (x_3 - \bar{x}_3)(y_3 - \bar{y}_3) = \sum x_3 y_3 - n \bar{x}_3 \bar{y}_3 = 26500.5 - 100(10.2)(24) = 2020.5$$

$$s_{yx4} = \frac{\sum (x_4 - \bar{x}_4)(y_4 - \bar{y}_4)}{n_4 - 1} = \frac{557.9}{59} = 9.46$$

$$\sum (x_4 - \bar{x}_4)(y_4 - \bar{y}_4) = \sum x_4 y_4 - n \bar{x}_4 \bar{y}_4 = 28097.9 - 60(15.3)(30) = 557.9$$

Şimdi varyansın tahmini değeri hesaplanabilir:

$$\begin{aligned} \hat{\sigma}^2(\bar{y}_{dra}) &= \sum_{h=1}^4 \frac{W_h^2}{n_h} (1 - f_h) (s_{yh}^2 - 2b_h s_{yhxh} + 2b_h^2 s_{xh}^2) \\ &= \frac{(0.10)^2 (0.996)}{40} (90 - 2(0.821)(26.93) + (0.821)^2 (32.8)) + \\ &\quad \frac{(0.50)^2 (0.996)}{200} (19.6 - 2(0.564)(9.19) + (0.564)^2 (16.29)) + \\ &\quad \frac{(0.25)^2 (0.996)}{100} (62.5 - 2(0.926)(20.41) + (0.926)^2 (22.04)) + \\ &\quad \frac{(0.15)^2 (0.996)}{60} (32.4 - 2(0.437)(9.46) + (0.437)^2 (21.64)) \\ &= 0.017 + 0.018 + 0.027 + 0.011 \\ &= 0.073 \end{aligned}$$

Anakütle ortalamasının tahminini içeren aralıklardan biri olma olasılığı 0.95 olan aralık ise aşağıdaki gibidir:

$$\hat{\bar{Y}}_{1,2} = \hat{\bar{Y}} \pm t\sigma_y$$

$$\hat{\bar{Y}}_{1,2} = 20.5719 \pm (1.96)(0.27)$$

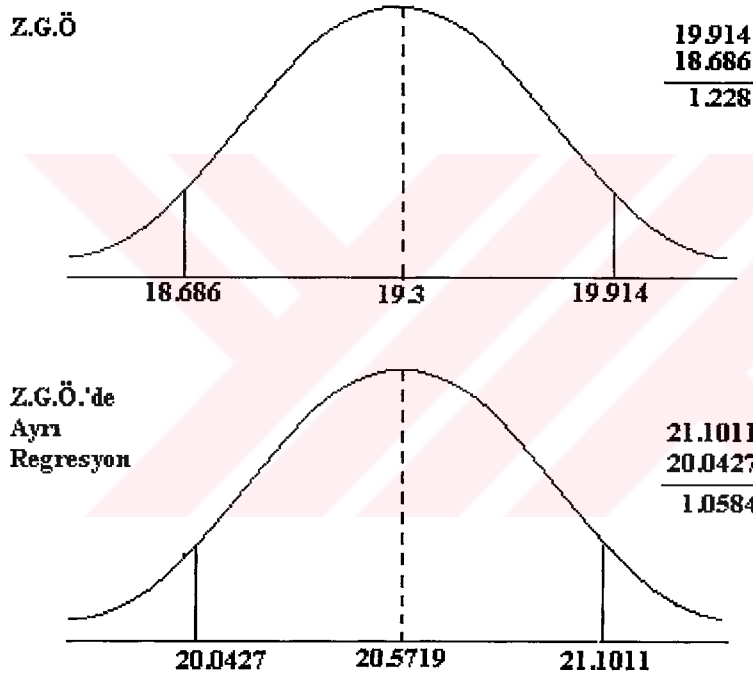
$$\hat{\bar{Y}}_1 = 20.0427 \text{ milyonTL / hafta}$$

$$\hat{\bar{Y}}_2 = 21.1011 \text{ milyonTL / hafta}$$

Tahminde etkinliđi arttırmak için önce ayrı regresyon yöntemi kullanılmış anakütle ortalamasına ilişkin tahmin ile bu tahminin varyansı bulunmuştur. Görüldüğü gibi bu aralık Z.G.Ö. yöntemine göre bulunan aralıktan daha dardır ve Şekil 2:'de bu iki aralığın karşılaştırılması daha iyi ifade edilmiştir.

Şekil 2:

**Z.G.Ö. Yöntemi İle Z.G.Ö.'de Ayrı Regresyon
Yöntemlerinin Karşılaştırılması**



7. Şimdi bileşik regresyon ile tahmin yöntemi kullanılacak ve bu yöntemin ayrı regresyon yöntemine göre üstünlüğü, yani varyansının daha düşük olduğu gösterilecektir.

Bileşik regresyon ile tahmin yönteminde her zümre için ayrı ayrı tahmin yapmak yerine, anakütle için tek bir tahmin denklemi kullanılacaktır. Bileşik regresyon ile tahminde tek bir b parametresi olacak ve bu parametre ayrı regresyon tahmininde hesaplanan b_{x1}, b_{x2}, b_{x3} ve b_{x4} parametrelerinin tartılı ortalaması olarak alınacaktır. Bileşik regresyon ile tahminde kullanılacak veriler Tablo:6'da sunulmuştur.

Tablo:6

Tahmin Verileri

W_h	$\bar{x}_h$	$\bar{y}_h$	$\bar{X}_h$
0.10	3.5	8	5
0.50	12.8	16	15
0.25	10.2	24	12
0.15	15.3	30	17

$$\bar{X} = \sum W_h \bar{X}_h = [(0.10 \times 5) + (0.50 \times 15) + (0.25 \times 12) + (0.15 \times 17)] = 13.55$$

Anakütle ortalamasının bileşik tahmini;

$$\hat{\bar{Y}}_{drb} = \bar{y}_b + b(\bar{X} - \bar{x}_b)$$

ile yapılır. Bu tahmin denkleminde gerekli olan hesaplamalar aşağıda sıra ile yapılmıştır.

$$\begin{aligned}\bar{y}_b &= \sum_{h=1}^4 W_h \bar{y}_h \\ &= [(0.10 \times 8) + (0.50 \times 16) + (0.25 \times 24) + (0.15 \times 30)] \\ &= 19.3\end{aligned}$$

$$\begin{aligned}\bar{x}_b &= \sum_{h=1}^4 W_h \bar{x}_h \\ &= [(0.10 \times 3.5) + (0.50 \times 12.8) + (0.25 \times 10.2) + (0.15 \times 15.3)] \\ &= 11.595\end{aligned}$$

$$\begin{aligned}b &= \sum W_h b_{xh} \\ &= [(0.10 \times 0.821) + (0.50 \times 0.564) + (0.25 \times 0.926) + (0.15 \times 0.437)] \\ &= 0.6612\end{aligned}$$

Şimdi anakütle ortalamasının bileşik tahminine geçilebilir:

$$\begin{aligned}\hat{\bar{Y}}_{drb} &= \bar{y}_b + b(\bar{X} - \bar{x}_b) \\ &= 19.3 + 0.6612 (13.55 - 11.595) \\ &= 20.593\end{aligned}$$

Bileşik regresyon tahmin yöntemi ile bulunan anakütle ortalamasının varyans tahmini ise;

$$\begin{aligned}\hat{\sigma}^2(\bar{y}_{drb}) &= \sum_1^4 \frac{W_h^2}{n_h} (1 - f_h) (s_{yh}^2 - 2bs_{yhx} + b^2 s_{xh}^2) \\ &= \frac{(0.10)^2 (0.996)}{40} [90 - 2(0.6612)(26.93) + (0.6612)^2 (32.8) \\ &\quad + \frac{(0.50)^2 (0.996)}{200} [19.6 - 2(0.6612)(9.19) + (0.6612)^2 (16.29)] \\ &\quad + \frac{(0.25)^2 (0.996)}{100} [62.5 - 2(0.6612)(20.41) + (0.6612)^2 (22.04) \\ &\quad + \frac{(0.15)^2 (0.996)}{60} [32.4 - 2(0.6612)(9.46) + (0.6612)^2 (21.64)] \\ &= 0.074\end{aligned}$$

Anakütle ortalamasının tahminini içeren aralıklardan biri olma olasılığı 0.95 olan aralık ise aşağıdaki gibidir:

$$\begin{aligned}\hat{\bar{Y}}_{1,2} &= \hat{\bar{Y}} \pm t\sigma_y \\ \hat{\bar{Y}}_{1,2} &= 20.593 \pm (1.96)(0.27) \\ \hat{\bar{Y}}_1 &= 21.1222 \text{ milyonTL / hafta} \\ \hat{\bar{Y}}_2 &= 20.0638 \text{ milyonTL / hafta}\end{aligned}$$

7. SONUÇ

Örnekleme yönteminde esas, tahminler bölünmesi varyansını küçük kılmaktır. Bunun iki yolu vardır; ya belli bir örnekleme yöntemi sınırları içinde kalıp tahmin yöntemini değiştirmek ya da belli bir tahmin yöntemini kullanarak örnekleme yöntemini değiştirmektir. Yaptığımız bu çalışmada örnekleme yöntemleri içinden Zümrelere Göre Örnekleme, tahmin yöntemlerinden de Regresyon Yöntemi ele alınarak uygulama yapılmıştır. Ayrıca regresyon yöntemi iki şekilde verilmiştir. Bunlar; ayrı ve bileşik regresyon tahmin yöntemleridir.

Ayrı regresyon yöntemi ile tahminde, ele alınan 4 zümre için 4 ayrı regresyon tahmin denklemi hesaplanmış ve bulunan bu 4 tahmin tartılandırılarak anakütle parametresi tahmin edilmiştir. Daha sonra bu tahminin varyansı bulunmuş ve anakütle parametresini içeren 0.95 olasılıklı geçerlilik hudutları bulunmuştur. 4 zümre için regresyon denkleminde bulunan tahminlerin sonuçları Tablo:7’de sunulmuştur.

Tablo:7

Tahmin Sonuçları

Zümre	$\hat{\bar{Y}}_h$
1	9.232
2	17.241
3	25.667
4	30.743

Yapılan bu tahminler tartılandırılarak anakütle ortalaması tahmin edilmiş ve bu değer $\hat{\bar{Y}}_{dra} = 20.5719$ çıkmıştır. Anakütle ortalamasının ayrı tahminine ilişkin varyansın tahmini değeri ise $\hat{\sigma}^2(\bar{y}_{dra}) = 0.073$ bulunmuştur. Anakütle ortalamasının ayrı tahminini içeren aralıklardan biri olma olasılığı 0.95 olan aralık ise $\hat{\bar{Y}}_1 = 20.0427$ milyon TL/hafta ve $\hat{\bar{Y}}_2 = 21.1011$ milyon TL/hafta olarak elde edilmiştir. Bu aralığın zümrelere göre örnekleme yöntemine göre bulunan $\hat{\bar{Y}}_1 = 19.914$ milyon TL/hafta ve $\hat{\bar{Y}}_2 = 18.686$ milyon

TL/hafta aralığında ve basit tesadüfi örnekleme yöntemine göre bulunan $\hat{Y}_1 = 20.182$ milyon TL/hafta ve $\hat{Y}_2 = 18.418$ milyon TL/hafta aralığından daha dar olduğu görülür. Bu ise ayrı regresyon tahmin yöntemini kullanmanın daha faydalı olduğunu ortaya koyar. Çünkü aralık ne kadar dar olursa, tahminlerin güvenilirliği de o derece artar.

Bileşik regresyon yöntemi ile tahminde ise tek bir regresyon tahmin denklemi kullanılmış ve anakütle parametresi buradan yola çıkarak tahmin edilmiştir. Bu tahmin denkleminde tek bir b parametresi kullanılacak ve bu parametre b_1, b_2, b_3 ve b_4 parametrelerinin tartılı ortalaması olarak hesaplanacaktır. Bileşik regresyon yöntemi ile tahminde anakütle ortalaması $\hat{Y}_{drb} = 20.593$ milyon TL/hafta olarak hesaplanmış ve anakütle ortalaması tahmininin varyans tahmini ise $\hat{\sigma}^2(\bar{y}_{drb}) = 0.074$ bulunmuştur. Anakütle ortalaması tahminini içeren aralıklardan biri olma olasılığı 0.95 olan geçerlilik hudutları ise $\hat{Y}_1 = 21.1222$ milyon TL/hafta ve $\hat{Y}_2 = 20.0638$ milyon TL/hafta olarak bulunmuştur. Bileşik regresyon ile tahminde b parametresi hesabında bir kıstas olmadığı için ayrı regresyon tahminine göre daha az etkindir ve bu özelliğinden dolayı daha az kullanılır.

Ayrı regresyon yöntemi ile tahmin yaparak bulunan anakütle ortalaması tahmininin varyansı bileşik regresyon yöntemi ile bulunan varyanstan daha küçüktür. Bu yüzden ayrı regresyon tahmin yöntemi uygulamada tercih sebebi olur.

KAYNAKLAR

1. Bağırkan,Ş. [1982], İstatistiksel Analiz, İstanbul: Önsöz Basım ve Yayıncılık.
2. Cochran,WW.G. [1977], Sampling Techniques, New York: John Wiley & Sons.
3. Çıngı,H. [1990], Örneklem Kuramı, Ankara : H.Ü. Basımevi
4. Hansen,M. , Madow,W.G. and Hurwitz,W.N. [1993], Sample Survey, New York: John Wiley & Sons.
5. Hahn,G.J., Meeker,W.Q. [1991], Statistical Intervals, New York: John Wiley & Sons.
6. İdil,O. [1980], Örneklem Teorisi ve İşletme Yönetiminde Uygulanması, İstanbul: Fatih Yayınevi Matbaası.
7. Lonard,W.R., Loftus,P.J. [1968], Sampling Theory, New York: Mc Graw Hill
8. Kish,L. [1965], Survey Sampling, New York: John Wiley & Sons.
9. Medhi,J. [1992], Statistical Methods, New York: John Wiley & Sons.
10. Moore,D.S. [1991], Statistics Concepts and Controversies, New York: W.H. Freeman & Company.
11. Orhunbilge, N. [1996], Uygulamalı Regresyon ve Korelasyon Analizi, İstanbul: Avcıol Basım Yayın.
12. Serper, Ö. [1992], Uygulamalı İstatistik, İstanbul: Filiz Kitabevi.
13. Serper,Ö., Aytaç, M., [1988], Örneklem, İstanbul: Filiz Kitabevi.
14. Sukhatme, P.V. [1958], Sampling Theory of Surveys with Applications, New York: The Iowa State Colage Press.
15. Thompson,S.K. [1992], Sampling, New York: John Wiley & Sons
16. Yoğurtçugil, M.K. [1976], Örneklem, İstanbul: Sermet Matbaası.