

137921

T.C.  
OSMANGAZİ ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ  
BİYOİSTATİSTİK ANABİLİM DALI

**ÇOK GRUPLU VERİ SETLERİNDE EKSİK  
GÖZLEM SORUNUNUN ÇÖZÜMLENMESİ VE  
SAĞLIK ALANINDA BİR UYGULAMA**

DOKTORA TEZİ

**TÜRKİYE YÜKSEKÖĞRETİM KURULU  
BİLİMSEL ARAŞTIRMA VE YATIRIM DİREKTÖRLÜĞÜ**

CENGİZ BAL

137921

DANIŞMAN

PROF. DR. KAZIM ÖZDAMAR

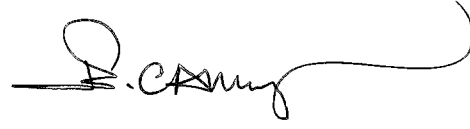
EYLÜL-2003

## KABUL VE ONAY SAYFASI

Cengiz BAL'ın DOKTORA TEZİ olarak hazırladığı “Çok Gruplu Veri Setlerinde Eksik Gözlem Sorununun Çözülmesi Ve Sağlık Alanında Bir Uygulama” başlıklı bu çalışma, jürimizce Doktora Öğretim Yönetmeliği'nin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

Tarih : 03.09.2003

ÜYE : Prof. Dr. Ersoy CANKÜYER



ÜYE : Prof. Dr. Kazım ÖZDAMAR (Danışman)



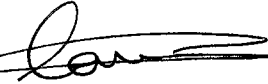
ÜYE : Yard. Doç. Dr. K. Setenay ÖNER



ÜYE : Yard. Doç. Dr. Fezan ŞAHİN

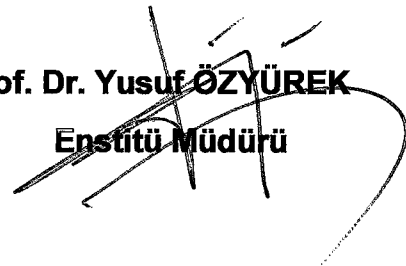


ÜYE : Yard. Doç. Dr. Canan DEMİRÜSTÜ



Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü Yönetim Kurulu'nun .....04.09.2003... tarih ve .....584...../.....1651..... sayılı kararı ile onaylanmıştır.

Prof. Dr. Yusuf ÖZYÜREK  
Enstitü Müdürü



# İÇİNDEKİLER

	<u>Sayfa</u>
İÇİNDEKİLER .....	I
ÖZET .....	V
SUMMARY .....	VI
ÇİZELGE DİZİNİ .....	VI
ŞEKİL DİZİNİ .....	XV
GÖSTERİM VE NOTASYONLAR .....	XVI
1- GİRİŞ .....	1
2- GENEL BİLGİLER .....	5
2.1- Eksik Veri Kavramı .....	7
2.2- Eksik Verinin Ortaya Çıkmasına Neden Olan Mekanizmalar.....	7
2.3- Eksik Veri Mekanizmaları .....	8
2.3.1- Tamamıyla Rassal Olarak Kayıp (TROK, MCAR) Mekanizması..	8
2.3.2- Rassal Olarak Kayıp (ROK, MAR) Mekanizması.....	9
2.3.3- İhmal Edilemez Kayıp (IEK, NI) Mekanizması.....	12
2.4- Eksik Verilerde Atama Yöntemleri.....	12
2.4.1-Tam Gözlemlerin Kullanılması Yöntemi (Listwise Method) .....	13
2.4.2- Eldeki Tüm Bilginin Kullanılması Yöntemi (Pairwise Method).....	14
2.4.3- Gözlemlerin Ve/Veya Değişkenlerin Silinmesi Yöntemi .....	17
2.5- Değer Atama Yöntemleri (Imputation Methods): .....	18

## İÇİNDEKİLER (Devam)

	<u>Sayfa</u>
2.5.1- Eksik Gözlem İle Tam Gözlemin Yer Değiřtirmesi (Case Substitution) .....	19
2.5.2- Eksik Gözlem İle Ortalamanın Yer Değiřtirmesi (Mean Substitution) .....	19
2.5.3- Hot Deck Atama (Hot Deck Imputation) Yöntemi .....	20
2.5.4- Cold Deck Atama (Cold Deck Imputation) Yöntemi .....	21
2.5.5- Regresyon Atama (Regression Imputation) Yöntemi .....	22
2.5.6- EM (Expectation-Maximization) Yöntemi .....	25
2.5.7- Çoklu Atama (Multiple Imputation-MI) .....	28
2.5.8- DA (Data Augmentation) Yöntemi .....	29
2.5.9 - Little'ın TROK (MCAR) Testi .....	31
3- GEREÇ VE YÖNTEMLER .....	33
3.1- Veri Türetimi .....	33
3.2- Eksik Veri Tamamlama Yöntemleri .....	34
3.2.1- Tam Gözlemlerin Kullanılması Yöntemi (Listwise Method) .....	35
3.2.2- Eldeki Tüm Bilginin Kullanılması (Karşılıklı Silme, Pairwise Method) .....	35
3.2.3- Değer Atama Yöntemleri (Imputation Methods): .....	36
3.2.3.1-Ortalama Atama Yöntemi (Mean Substitution) .....	36
3.2.3.2- Regresyon Atama (Regression Imputation) Yöntemi .....	37

## İÇİNDEKİLER (Devam)

	<u>Sayfa</u>
3.2.3.3- EM (Expectation-Maximization) Yöntemi .....	38
3.3- Little'ın TROK (MCAR) Testi .....	40
3.4- Eksik Veri Analizinde Belirtici İstatistikler .....	42
4-BULGULAR .....	43
4.1- 50 Birimlik Türetilmiş Veri Setleri .....	44
4.1.1- Düşük Korelasyonlu 50 Birimlik Veri Seti .....	44
4.1.2- Yüksek Korelasyonlu 50 Birimlik Veri Seti .....	50
4.2- 100 Birimlik Türetilmiş Veri Setleri .....	55
4.2.1- Düşük Korelasyonlu 100 Birimlik Veri Seti .....	55
4.2.2- Yüksek Korelasyonlu 100 Birimlik Veri Seti ....	60
4.3- 200 Birimlik Türetilmiş Veri Setleri .....	65
4.3.1- Düşük Korelasyonlu 200 Birimlik Veri Seti ..	66
4.3.2- Yüksek Korelasyonlu 200 Birimlik Veri Seti .....	71
4.4- 300 Birimlik Türetilmiş Veri Setleri .....	77
4.4.1- Düşük Korelasyonlu 300 Birimlik Veri Seti .....	77
4.4.2- Yüksek Korelasyonlu 300 Birimlik Veri Seti .....	82
4.5- 400 Birimlik Türetilmiş Veri Setleri .....	87
4.5.1- Düşük Korelasyonlu 400 Birimlik Veri Seti .....	88
4.5.2- Yüksek Korelasyonlu 400 Birimlik Veri Seti .....	93

## İÇİNDEKİLER (Devam)

	<u>Sayfa</u>
4.6- 500 Birimlik Türetilmiş Veri Setleri .....	98
4.6.1- Düşük Korelasyonlu 500 Birimlik Veri Seti .....	98
4.6.2- Yüksek Korelasyonlu 500 Birimlik Veri Seti .....	104
5- TARTIŞMA VE SONUÇ .....	110
6- BİR UYGULAMA .....	120
KAYNAKLAR DİZİNİ .....	125
ÖZGEÇMİŞ .....	134

## ÖZET

Eksik veri problemi, yıllardır arařtırmacıların bilimsel bilgiyi üretmelerinde önlerindeki en büyük engellerden biri olmuřtur. Özellikle sađlık alanında toplanan verilerde eksik gözlem deđerleri, analiz aşamasında ve sonuçların genellenmesi aşamasında karar verirken, arařtırmacıya zor anlar yaşatmakta ve hatalı kararlar vermesine neden olabilmektedir. Eksik veriler, gerek saha arařtırmalarında gerekse laboratuvar koşullarında yapılan çalışmalarda arařtırmacının karşısına çıkabilmektedir. Arařtırmacılar öncelikle eksik veriyi ortaya çıkartan nedenleri ve eksik verinin arařtırma içerisindeki önem derecesini belirlemelidirler.

Bu arařtırma bir benzetim çalışmasıdır ve farklı birim sayılarına sahip türetilmiş veri setleri yardımıyla, eksik veri atama yöntemleri karşılaştırılmıştır. Türetilen veri setleri, sırasıyla; 50, 100, 200, 300, 400 ve 500 birim içerecek şekilde, çok deđişkenli standartlaştırılmış normal dağılım varsayımından yararlanılarak ve düşük ve yüksek korelasyonlu deđişkenleri içerecek şekilde türetilmiştir. Rasgele koşullar altında %5, %10, %15 ve %20 eksiltelen veri setleri TROK (MCAR) yapıya sahiptir. Elde edilen veri setlerinin analizleri sonucunda Listwise, Pairwise, Ortalama, Regresyon ve EM atama yöntemleri düşük hacimli örneklerde tutarsız sonuçlar vermişlerdir ( $n < 200$ ). Fakat tam veri setine en yakın sonuçlar, EM atama yönteminden elde edilmiştir. Benzer şekilde, yüksek hacimli örneklerden elde edilen sonuçlarda da, EM atama yöntemi tam veri setlerine oldukça yakın sonuçlar vermiştir. EM atama yöntemi özellikle 200 birim ve üzerindeki veri setlerinde, %5-%20 arasındaki eksik veri yapılarında diđer atama yöntemlerine göre daha iyi sonuçlar vermiştir.

Türetilmiş veri setlerinden elde edilen sonuçlara göre, bazı deđişkenlerinde yaklaşık %20 gözlem deđeri eksik olan sađlık alanındaki gerçek bir veri seti EM atama yöntemi ile tamamlanmış ve analizler eksik veri seti ile karşılařtırmalı olarak sunulmuřtur.

**Anahtar Kelimeler:** Atama yöntemleri, EM, Regresyon atama, Listwise, Pairwise, Ortalama atama.

## SUMMARY

The problem of missing values has been a big difficulty to produce scientific information. Especially missing observations create difficulties for the analysis and for the generalization of the results in health sciences and could make one give wrong decisions. Missing values can come up in either survey research or in laboratory conditions. Researchers, first of all, should investigate the reasons for missing value and the importance of missing values in data set.

In this thesis, simulation study is done, using simulated data sets having different sample sizes, and also some imputation methods are compared. Simulated data is taken from multivariate Normal distribution in which either high correlation matrix or low correlation matrix is assumed, and the sample sizes are set to 50, 100, 200, 300, 400, and 500. 5%, 10%, 15%, and 20% of simulated samples are randomly set missing, and MCAR (TROC) property is observed on the reduced samples. It is seen that Listwise, Pairwise, Mean, Regression, and EM imputation methods have consistency problems for small sample sizes ( $n < 200$ ). However, the closest results to the results of full data set are produced by EM method. Similarly, for large sample sizes, EM produced the closest results to the full data set. Also, EM imputation produced better results for large samples ( $n > 200$ ) with missing value percentages between 5-20.

Based on the information gained by the simulation study, a case data set on which some variables having 20% missing value is completed using EM imputation method, and the results are given in detail.

**Key Words :** Imputation Methods, EM, Regression, Listwise, Pairwise, Mean Imputation.



## ÇİZELGE DİZİNİ

	<u>Sayfa</u>
Çizelge 4.1: Türetilen veri setlerine ait Little'nin MCAR test istatistiği sonuçları .....	43
Çizelge 4.2- 50 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri .....	44
Çizelge 4.3- 50 birim içeren düşük korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri .....	45
Çizelge 4.4- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri .....	45
Çizelge 4.5- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları .....	46
Çizelge 4.6- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	47
Çizelge 4.7- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	48
Çizelge 4.8- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	48
Çizelge 4.9- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	49
Çizelge 4.10- 50 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri .....	50
Çizelge 4.11- 50 birim içeren yüksek korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri .....	50
Çizelge 4.12- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri	51
Çizelge 4.13- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları .....	52
Çizelge 4.14- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	53

## ÇİZELGE DİZİNİ ( Devam)

	<u>Sayfa</u>
Çizelge 4.15- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	53
Çizelge 4.16- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	54
Çizelge 4.17- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	54
Çizelge 4.18- 100 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri .....	55
Çizelge 4.19- 100 birim içeren düşük korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri .....	56
Çizelge 4.20- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri .....	56
Çizelge 4.21- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları .....	57
Çizelge 4.22- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	58
Çizelge 4.23- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	59
Çizelge 4.24- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	59
Çizelge 4.25- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	60
Çizelge 4.26- 100 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri .....	61
Çizelge 4.27- 100 birim içeren yüksek korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri .....	61

## ÇİZELGE DİZİNİ ( Devam)

	<u>Sayfa</u>
Çizelge 4.28- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri .....	62
Çizelge 4.29- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları .....	62
Çizelge 4.30- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	63
Çizelge 4.31- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	64
Çizelge 4.32- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	64
Çizelge 4.33- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	65
Çizelge 4.34- 100 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri .....	66
Çizelge 4.35- 200 birim içeren düşük korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri .....	66
Çizelge 4.36- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri .....	67
Çizelge 4.37- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları .....	68
Çizelge 4.38- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	69
Çizelge 4.39- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	69
Çizelge 4.40- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	70

## ÇİZELGE DİZİNİ ( Devam)

	<u>Sayfa</u>
Çizelge 4.41- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	70
Çizelge 4.42- 100 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri .....	71
Çizelge 4.43- 200 birim içeren yüksek korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri .....	72
Çizelge 4.44- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri .....	72
Çizelge 4.45- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları .....	73
Çizelge 4.46- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	74
Çizelge 4.47- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	75
Çizelge 4.48- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	75
Çizelge 4.49- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	76
Çizelge 4.50- 300 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri .....	77
Çizelge 4.51- 300 birim içeren düşük korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri .....	77
Çizelge 4.52- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri .....	78
Çizelge 4.53- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları .....	79
Çizelge 4.54- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	80

## ÇİZELGE DİZİNİ ( Devam)

	<u>Sayfa</u>
Çizelge 4.55- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	80
Çizelge 4.56- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	81
Çizelge 4.57- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	81
Çizelge 4.58- 300 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri .....	82
Çizelge 4.59- 300 birim içeren yüksek korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri .....	83
Çizelge 4.60- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri .....	83
Çizelge 4.61- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları .....	84
Çizelge 4.62- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	85
Çizelge 4.63- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	86
Çizelge 4.64- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	86
Çizelge 4.65- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	87
Çizelge 4.66- 400 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri .....	88
Çizelge 4.67- 400 birim içeren düşük korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri .....	88

## ÇİZELGE DİZİNİ ( Devam)

	<u>Sayfa</u>
Çizelge 4.68- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri .....	89
Çizelge 4.69- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları .....	90
Çizelge 4.70- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	91
Çizelge 4.71- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	91
Çizelge 4.72- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	92
Çizelge 4.73- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	92
Çizelge 4.74- 400 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri .....	93
Çizelge 4.75- 400 birim içeren yüksek korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri .....	93
Çizelge 4.76- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri .....	94
Çizelge 4.77- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları .....	95
Çizelge 4.78- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	96
Çizelge 4.79- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	97
Çizelge 4.80- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	97

## ÇİZELGE DİZİNİ ( Devam)

	<u>Sayfa</u>
Çizelge 4.81- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	98
Çizelge 4.82- 500 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri .....	99
Çizelge 4.83- 500 birim içeren düşük korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri .....	99
Çizelge 4.84- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri .....	100
Çizelge 4.85- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları .....	100
Çizelge 4.86- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	102
Çizelge 4.87- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	102
Çizelge 4.88- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	103
Çizelge 4.89- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	103
Çizelge 4.90- 500 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri .....	104
Çizelge 4.91- 500 birim içeren yüksek korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri .....	104
Çizelge 4.92- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri .....	105
Çizelge 4.93- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları .....	106
Çizelge 4.94- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	107

## ÇİZELGE DİZİNİ ( Devam)

	<u>Sayfa</u>
Çizelge 4.95- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	107
Çizelge 4.96- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	108
Çizelge 4.97- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları .....	108
Çizelge 6.1: Değişkenlere ilişkin geçerli ve eksik gözlem sayısı .....	121
Çizelge 6.2: Toplam 214 kadına ait değişkenler, alt kategorileri, sayı ve yüzdeleri .....	121
Çizelge 6.3- Hastalardan elde edilen değişkenlerin eksik olma sayı ve yüzde değerleri .....	122
Çizelge 6.4: Veri setine ilişkin eksik olma paterni .....	122
Çizelge 6.5: Değişkenlere ilişkin ortalama ve standart sapma değerleri ....	122
Çizelge 6.6: Eksik ve EM atama yöntemi ile tamamlanan veri setlerinde yapılan t testleri .....	123



## ŞEKİL DİZİNİ

	<u>Sayfa</u>
Şekil 2.1: Eksik veri yapıları .....	10
Şekil 2.2: <b>Önce:</b> Eksik veri yapısına ait dağılım grafiği, <b>Sonra:</b> Eksik gözlem değerlerinin yerine ortalamanın atıldığı durumda veri yapısına ait dağılım grafiği .....	20



## GÖSTERİM VE NOTASYONLAR

Bu tezde kullanılan gösterim ve notasyonlar aşağıda kısaca tanımlanmıştır.

$X$	: Veri matrisi
$x_{ij}$	: i. birimin, j. değişken değeri
$v$	: değişken sayısı
$n$	: birim sayısı
$n_i$	: i. değişkendeki eksik olmayan değer sayısı
$n_{ij}$	: i. ve j. değişkenlerin eksik olmayan çift sayısı
$n_c$	: Tamamlanmış (complete) veri sayısı
$n_E$	: Cevapları tam olan birimlerin sayısı
$J$	: Değişken indeksi
$J_{\#}=J(\text{koşul})$	: 'koşulu' sağlayan değişken indeksi
$I$	: Birim indeksi
$I(k_1, \dots, k_l)$	: değişkenleri $(k_1, \dots, k_l)$ eksik olmayan birim indeksi
$I(J)$	: Tamamlanmış (eksik olmayan, complete) birim indeksi
$\mathbf{a}=[a_i]$	: i. elemanı $a_i$ olan vektör
$\mathbf{A}=[a_{ij}]$	: i. satır, j. sütun elemanı $a_{ij}$ olan matris

## 1- GİRİŞ

Günümüzde, teorik ve uygulamalı bilim alanlarında bilimsel arařtırmalara büyük önem verilmektedir. Bilimsel arařtırmalarda; planlama, arařtırma alanlarının ve örneklemin seçilmesi, veri toplama, verilerin tablo ve grafiklerle özetlenmesi, hipotezlerin oluşturulması, hipotezlerin test edilmesi, sonuçların yorumlanması ve genellenmesi, bilimsel raporların hazırlanması ve sunum teknikleri ile ilgili bilimsel bilgi üretiminde tek tip yaklaşımlar kullanmak büyük önem taşımaktadır. Bilimsel arařtırmalarda bilgi iletişimini kolaylařtıracak ortak yaklaşımlar kullanmak ve etkin çalışmalar yapabilmek için istatistiksel yöntemlerden yararlanmak gerekmektedir (29).

Arařtırmalar, toplumdaki birimlerin tümü ya da bir bölümünün bir veya birden fazla deęişkenini uluslararası bilimsel ve teknik standartlara uygun bir biçimde ölçerek, tartarak ya da sayarak, sayısal deęerlerini saptamayı ve bu sayısal deęerleri kullanarak, toplumda bir deęişkenin ana eğilimini belirlemeyi amaçlar. Ayrıca arařtırmalar, birimlerin çoğunun toplandığı ortak noktaları ve genel eğilimleri belirleyerek toplum için geçerli ve güvenilir kararlara ulaşmak, genellemeler yapmak ve yasalar ortaya koymak amacıyla yapılan çalışmalardır (28).

Arařtırmalardan geçerli ve güvenilir bilimsel bilgi elde etmek için, çalışmaların belirli bir plan, program ve uygulama planına (protokol) göre yapılması gerekir. Bilimsel bilgi üretecek özellikte arařtırma yapabilmek için arařtırma planlamasını doğru yapmak gerekir. Arařtırma planlaması, arařtırmadan geçerli ve güvenilir veri elde etmek için izlenmesi gereken en önemli aşamalardan biridir (18, 28, 31). Elde edilen veri setinin analize hazır hale getirilmesi aşamasında dikkat edilmesi gereken birtakım unsurlar bulunmaktadır. Bunlar verilerdeki hata kontrolüdür. Özellikle veri setinde hatalı giriş, min-max, aşırı deęer (outlier) ve ya eksik veri gibi, arařtırmanın sonucunu direk etkileyecek ve hatalara sebebiyet verecek kontrollerinin yapılması gerekmektedir (29).

Arařtırmacının dođru kararlar verebilmesi ve uygulamalarda bařarılı olabilmesi iin iyi nitelikli bir enformasyona sahip olması gereklidir. Nitelikli bir enformasyon nitelikli bir veri ile elde edilebilir. İyi nitelikli bir veri; dođru, tam, gvenilir, yararlı, kullanılabilir, ekonomik, basit, gncel, zamanında ve denetlenebilir zelliklere sahip olmalıdır (28, 43, 44). İyi tasarlanmayan ve toplanan verileri bu kořulları sađlamayan arařtırmalar, hibir istatistiksel yntemle dzeltilemeyecek problemler oluřturulabilir. Tıbbi olarak mdahale yapılmasının mmkn olmadığı bir hastaya verilen ilalar fayda sađlamadıđı gibi, iyi tasarlanmamıř bir arařtırmaya uygulanan istatistiksel analiz de fayda sađlamaz. Bu bakımdan, bir arařtırma iin uygun bir alıřma dzeninin seimi yapılırken, arařtırılacak konunun zellikleri, zaman, maddi kaynaklar ve personel gibi unsurlar da dikkate alınarak arařtırmanın hatasız planlanması gerekir. Aksi takdirde birimlerin ve deđiřkenlerin tanımlanmasında, verilerin toplanmasında hatalar, eksiklikler gzlenmekte ve kaynakların israfı yanında elde edilen sonuların da gvenilirliđi azalmaktadır (1, 5, 11).

Eksik veri problemi, yıllardır arařtırmacıları meřgul etmiřtir. ok iyi planlanmış bir arařtırmada bile eksik verilere sık sık rastlanmaktadır. Eksik veriler gerek saha arařtırmalarında gerekse laboratuvar kořullarında yapılan alıřmalarda arařtırmacının karřısına ıkabilir. Arařtırmacılar ncelikle eksik veriyi ortaya ıkartan nedenleri ve eksik verinin arařtırma ierisindeki nem derecesini belirleyebilmelidir (2, 7, 10).

Tekdeđiřkenli istatistiksel analiz yntemleri, incelenen deđiřken zerinde etkide bulunan i ve dıř faktrleri toplumdaki tm birimler iin sabit ya da trdeř kabul ederek zmlenmelere gitmeyi amalamaktadır. Trdeř ya da sabit kořulları sađlamak ođu zaman mmkn deđildir. Bu nedenle bir problemin zmnde ok sayıda deđiřkeni dikkate alarak zmler retmek gereki bir yaklařımdır. okdeđiřkenli istatistiksel analiz ise incelenen olay evresindeki ok sayıda isel ve dıřsal (endojen ve ekzojen) faktrleri dikkate alarak, problemi dođasındaki yapısına iliřkin bilgilere gre (gerek oluřumuna gre) incelemek ve zmlere ulařmak iin geliřtirilmiř yntemler bttndr (30).

Eksik veriler, yapılacak olan istatistiksel analizlerde önemli problemler yaratmaktadır. Çünkü istatistiksel analizler ve bu analizlerin yapılmasına olanak veren ilgili paket programlar, verilerin tümünün var olduğu (tam, complete) durumlar için geliştirilmiştir. Araştırmalarda yer alan ve eksik veri ya da verileri içeren birimler, bir bilgi yokluğunu temsil ederler, dolayısıyla bir bilgi kaybına neden olurlar. Standart istatistiksel yöntemler ve bunların çözümlerini sağlayan ve kolaylaştıran paket programlar, tam bilgi durumu için düzenlenmişlerdir. Bu analizler, eksik veri içeren veri setlerine uygulandıklarında genellikle eksik veriye sahip olan gözlemler işlem dışı bırakılır, dolayısıyla yanlış tahminlere neden olabilmektedir. Ayrıca inceleme kapsamındaki veri setinde, bir ya da daha fazla değişkenin eksik veri içermesi nedeniyle ilgili birimlere ait bilgilerin analizden çıkarılması hesaplanan istatistiklerin kararlılığını düşürmekte ve çalışmanın geçerliğini ve genellenebilirliğini etkilemektedir. Araştırmalardaki eksik veri problemi için, çeşitli çözüm ve atama yöntemleri geliştirilmiştir. Genel olarak, eksik veri problemi ile karşılaşıldığında yapılabilecek işlemler veri setindeki eksik gözlemleri atmak, yeni gözlemler eklemek ya da veri setindeki eksik değerleri çeşitli istatistiksel yöntemlerle gidermeye çalışmak şeklinde incelenir. Bu yöntemler, geçerli ve tutarlı sonuçlara ulaşabilmek açısından oldukça önemlidir (7, 22-27, 35, 36, 46).

Eksik veri çalışmaları, 1970'lerde ad-hoc prosedürleri, gözlem birimlerinin atılması ve tekli atama işlemleri üzerinde çalışmalar yapılarak başlamıştır. İlerleyen yıllarda, özellikle 1980'li yıllara gelindiğinde, dikkat daha çok benzerliğe dayalı tahmin prosedürleri ile EM (Expectation Maximization) algoritmasında yoğunlaşmıştır. 1990'lı yıllara gelindiğinde ise çoklu atama (Multiple Imputation), MCMC (Markov Chain Monte Carlo) teknikleri ve Bayesçi teknikler üzerinde yoğunlaşmalar görülmektedir. Bu tarihçe içerisinde en çok dikkat edilmesi gereken nokta ise eksik veriye bakışta meydana gelen değişimlerdir. 1980'li yıllardan itibaren eksik veriler araştırmalarda bir sorun olmaktan yavaş yavaş çıkmaya başlamıştır. Eksik veri içeren birimlerin çalışmadan çıkarılması yerine bu birimlerin analize alınması için uygun olan istatistiksel yöntemlerin kullanılması önerilmiştir. Araştırmadaki birim sayısına, araştırma tipine ve elde edilen veri tipine uygun olabilecek değişik yöntemlerin

geliştirilmesine yönelik çalışmalar yoğun biçimde devam etmektedir (16, 24-26).

Bu çalışmada;

Eksik gözlem içeren tek ve çok gruplu, tek ve çok değişkenli veri setlerinde eksik veri probleminin giderilmesine yönelik istatistiksel yöntemler incelenmiştir.

Koşullu türetilen veri setlerinde eksik veri problemlerinin çözümüne yönelik farklı atama yöntemlerinin etkinlikleri incelenmiştir.

Normal dağılım gösteren değişkenlere sahip, farklı birim sayıları içeren; tam, eksiltilmiş ve atama yöntemleri ile tamamlanmış veri setlerinden elde edilen analiz sonuçları karşılaştırılmıştır.

Değişik koşullara göre türetilmiş veri (conditionally generated data) setlerinde bu yöntemlerin parametre tahminlerine etkileri araştırılmıştır.

Bu araştırmada yöntemlerin karşılaştırılmasında yararlanılan koşullu türetilmiş veri setleri;

- Çokdeğişkenli normal dağılıma sahip veri setleri,
- Farklı örnek hacimlerine sahip veri setleri ( $n=50, 100, 200, 300, 400, 500$ ),
- Değişkenlerindeki eksik veri sayısı farklı olan veri setleri (%5, %10, %15, %20),
- Farklı tip ve özelliklere sahip gruplara ilişkin veri setleri,
- Değişkenler arasındaki korelasyonların yüksek ya da düşük olduğu veri setleri,

olarak ele alınmıştır.

Bu analizler sonucunda tutarlılığı ve etkinliği saptanan atama yöntemi sağlık alanından elde edilen gerçek bir veri setine uygulanmıştır.

## 2- GENEL BİLGİLER

Araştırmalarda yararlanılan veri toplama yöntemlerinin çoğunda, gerek araştırmacıdan, gerek araştırılan birimlerden, gerek araştırmanın planından ya da kontrol altına alınamayan nedenlerden dolayı eksik verilerle karşılaşılabilir. Araştırmacı ne kadar dikkatli, titiz, tam olarak araştırmanın gereğini yerine getirirse getirsin elde olmayan veya engellenemeyen nedenlerden dolayı eksik veriler oluşmaktadır. Asıl problem eksik verinin nasıl oluştuğu değil, eksik veri ile nasıl başa çıkılacağıdır (7, 22, 39).

Araştırmacılar, yaptıkları bilimsel çalışmalarda, veri toplama aşamasında ne kadar dikkatli olsalar da, eksik veri problemleriyle sıkça karşılaşmaktadırlar. Bu eksik veriler istatistiksel analizlerin veri setine uygulanması aşamasında araştırmacıya sorun olmaktadır (2, 10, 27). Çünkü mevcut istatistiksel yöntemler, bunların uygulama aşamasında kullanılan ve araştırmacıya oldukça kolaylık sağlayan istatistiksel yazılımlar, veri setindeki gözlemlerin tümünün tam (eksiksiz, complete, non-missing) olduğu durumlar için geliştirilmişlerdir. Elde edilen veri setinde eksik gözlemler (eksik veriler) olduğunda, eldeki veri setinin analizi aşamasında büyük zorluklar yaşanmaktadır (35). Bu noktadan hareketle araştırmacılar, eksik veri analizi yöntemlerini geliştirmişlerdir. Eksik veri analizine (Missing value analysis), eksik veri içeren herhangi bir veri seti için istatistiksel analiz uygulamadan önce başvurulması gereken bir ön analiz olarak bakılmalıdır. Eksik veri analizi ile, veri setindeki eksik veriler, uygun değerlerle tamamlanır ve tamamlanmış veri seti ile istatistiksel işlemler yapılır ve sonuca gidilir (24, 38).

Araştırmalarda öncelik eksik verinin ortaya çıkmasını önlemektir. Ancak tüm birimlerin tüm değişkenlerine ilişkin verileri toplamak bazı durumlarda mümkün olmamaktadır. Eksik verinin söz konusu olduğu durumlarda öncelikle eksik veriyi ortaya çıkartan nedenleri ve eksik veri setinde eksik veri içeren birimlerin sayısı, genel birim içindeki yüzdesi, değişkenin dağılımı, değişkenler arasındaki korelasyon düzeyi gibi bilgileri elde etmek gerekmektedir. Veri matrisi hakkındaki bu bilgiler uygulanacak

eksik veri analizi yöntemlerini belirlemede önemli rol oynamaktadır (2).

Araştırmalarda yer alan eksik veri içeren birimler, bir bilgi eksikliğini temsil ederler. Dolayısıyla da bir bilgi kaybına neden olurlar. Standart istatistiksel yöntemler ve bunların uygulanmasında kullanılan yazılımlar tam veri setleri için düzenlendiklerinden, eksik veri setlerine uygulandıklarında yanlış tahminlere neden olurlar (7, 22, 27).

Gözlem birimi kaynaklı (sorulara yanıt vermeme, bulunamama vb.) ya da gözlem birimi dışında (veri toplama, kayıt ve veri girişi hataları) nedenlerden kaynaklanan ve eksik veriye yol açan sürece, eksik veri süreci denir. Bazen, eksik veri süreci ile karşılaşılabilen önceden tahmin edilebilir ve araştırma planında gerekli önlemler alınabilir. Ancak, özellikle gözlem birimi kaynaklı olarak ortaya çıkan eksik veri yapıları, önceden öngörülemezdir. Bu durumda, veride eksik değerleri ortaya çıkaran bir yapı olup olmadığının araştırmacı tarafından incelenmesi gerekir (2, 24, 36).

Bu incelemede genellikle;

1. Eksik verilerin gözlemlere rasgele dağılıp dağılmadığı ve belirgin bir yapı oluşturup oluşturmadıkları,
2. Eksik verinin ne kadar olduğu (ne sıklıkta ortaya çıktığı) sorularına yanıt aranır.

Eksik veri ile ilgili yapılar söz konusu ve eksik verinin araştırma içerisindeki sayısı herhangi bir önlem almayı gerektirecek düzeyde ise, veride eksik veri sürecinin varlığından söz edilebilir. Bu durumda, eksik veriler ile elde edilen istatistiksel sonuçlar, çalışmada kullanılan değişkenlerin eksik veri sürecinden etkilendiği derecede yanlış olur. Fakat, eksik verinin araştırma sonuçlarına etkisi, sonuçlar üzerindeki yanlışlığı ile sınırlı değildir. Örneğin, eksik veriye ilişkin problemlerin giderilemediği çalışmalarda, eksik verili gözlemlerin veriden çıkartılması söz konusu olur. Bu durum ise, gözlem sayısında ciddi bir azalmayı beraberinde getirir ve araştırmanın başında, yeterli olarak düşünülen bir örnek, yetersiz sayıdaki bir örneğe dönüşür. İncelenen veri setinde bir ya da birden çok değişkenin eksik veri içermesi nedeniyle ilgili gözlem birimlerinin veri setinden çıkartılması, hesaplanan istatistiklerin kararlılığını azaltırken çalışmanın bilimsel olarak geçerliğini ve genellenebilirliğini etkileyebilir (2, 9, 22, 38).



Arařtırmalarda eksik veri problemi ile karřılařıldığında, çoęunlukla veri setinde bulunan eksik gözlemlere ait birimler silinir, yeni gözlemler ilave edilir ya da verideki eksik deęerleri çeřitli atama yöntemleri yardımıyla tahmin edilmeye çalışılarak, veri seti istatistiksel analizlerin yapılması için uygun hale getirilir (35).

### **2.1- Eksik Veri Kavramı**

Arařtırmalarda, eksik veri problemlerini çözmeye yönelik yaklařımlar öncesinde, eksik veri yapılarının ve bazı tanımlayıcı istatistiklerin incelenmesi gereklidir. Bu inceleme ile eksik verilerin veri setinde nerelerde ve ne miktarda yoğunlařtığı, bir deęişken için eksik olan bir deęerin dięer deęişken ya da deęişkenler için eksik olma eğiliminin de olup olmadığı, eksik veri yapısının dięer deęişkenlerin deęerleri ile ilişkili olup olmadığı, deęişken silindiğinde analize alınacak gözlem sayısı ve deęişkenlerin daęılım özellikleri belirlenmeye çalışılır. Veri setindeki eksik deęerlerin giderilmesi öncesinde ařağıdaki adımların takip edilmesi gerekmektedir (2).

- Eksik veri çeřitleri
- Eksik veri yapısının incelenmesi
- Eksik veri sürecinde rasgelelięin arařtırılması
- Eksik veri sorununun giderilmesine yönelik yaklařımlar

Eksik veri problemi için çeřitli çözüm ve atama yöntemleri geliřtirilmiřtir. Bu çözüm ve atama yöntemleri öncesinde eksik veri mekanizmalarının açıklanması gerekir. Çünkü, veri setine hangi çözüm veya atama yönteminin uygulanacaęı, dięer bir anlatımla hangi çözüm ve atama yönteminin uygun olabileceęi, eksik veri mekanizmalarına baęlıdır (24).

### **2.2- Eksik Verinin Ortaya Çıkmasına Neden Olan Mekanizmalar**

Eksik veri yapılarındaki eksik veri mekanizmaları hangi atama yönteminin/yöntemlerinin seçilebileceęini belirleme bakımından oldukça önemlidir. Çünkü,

eksik veri atama yöntemlerinin özellikleri, bu mekanizmalar içerisindeki bağımlılığın doğasına bağlıdır. Eksik verilerin analizinde, mekanizmanın rolü Rubin'in (1976) teorisinde formüle edilinceye kadar araştırmacılar tarafından ihmal edilmiştir. Rubin'in teorisinde, eksik veri göstergeleri rassal değişkenler olarak işlem görmüş ve bir dağılıma atanmıştır (24).

### **2.3- Eksik Veri Mekanizmaları**

Eksik veriler için çözüm ve atama (yerine koyma) yöntemlerinin doğru kullanımı, geçerli sonuçlara ulaşabilmek açısından oldukça önemlidir. Uygun çözüm ve atama yönteminin seçimi, eksik veri mekanizmalarına bağlıdır. Little ve Rubin bu mekanizmaları üç temel kategoriye ayırmıştır. Bunlar; Tamamıyla Rassal Olarak Kayıp (TROK) (Missing Completely at Random, MCAR), Rassal Olarak Kayıp (ROK) (Missing at Random, MAR) ve İhmal Edilemez Kayıp (IEK), (Nonignorable, NI) olarak isimlendirilmiştir (3, 6, 12, 13, 16)

#### **2.3.1- Tamamıyla Rassal Olarak Kayıp (TROK, MCAR) Mekanizması**

A ve B gibi iki değişken ele alındığında, eğer cevap olasılığı A ve B değişkenlerinden bağımsız ise, kayıp veri TROK olarak ele alınır. Diğer bir anlatımla 'kayıp olma (missingness)' analizde yer alan spesifik (özel) değişkenlerle ilişkili değildir (2, 15, 22, 39).

TROK koşullarının sağlanıp sağlanmadığı yanıt verenler ile yanıt vermeyenler arasındaki gözlenen verilerin dağılımlarının karşılaştırılmasıyla belirlenebilir. Eğer, veriler için TROK sağlanıyorsa, liste bazında silme (listwise deletion, tam gözlemlerin kullanılması) veya diğer bir ifade ile birim bazında silme (casewise deletion) çözüm yöntemleri iyi bir seçim oluşturacaktır. Bu yöntemler pek çok istatistiksel paket programında mevcuttur. Avantajı ise basitliği ve hesaplama süresinin kısalığıdır. Eğer veriler TROK özelliği taşımiyorsa, sonuçlar yanlı olmakta, bu durumda robust yöntemler çözümlene için uygun olmaktadır (24, 39).

$Y=(y_{ij})$ ,  $(n \times K)$  boyutunda, eksik değer içermeyen dikdörtgen bir veri setini gösterebilir. Burada;  $i$ . satır için  $y_i=(y_{i1}, \dots, y_{ik})$  şeklinde gösterilir ve  $y_{ij}$ ,  $Y_j$  değişkeninin  $i$ . biriminin değeridir. Eksik veri yapısı içeren veri setlerinde, eksik veri gösterge matrisi  $M=(m_{ij})$  olarak tanımlanır. Burada; eğer  $y_{ij}$  eksik veri ise  $m_{ij}=1$  değerini, değilse 0 değerini alır.  $M$  matrisi eksik veri yapısını tanımlar. Şekil 2.1'de eksik veri yapısı örnekleri gösterilmiştir. Bazı yöntemler, herhangi bir eksik veri yapısına uygulanmak için ele alınırken, bazı yöntemler ise özel yapılara uygulanmak için kısıtlanmıştır (24).

Yukarıda tanımlandığı gibi;  $Y=(y_{ij})$  veri setini,  $M=(m_{ij})$  eksik veri gösterge matrisini gösterebilir. Eksik veri mekanizması,  $Y$  verilmiş iken  $M$ 'nin koşullu dağılımıdır ve  $f(M | Y, \phi)$  şeklinde gösterilir. Burada  $\phi$  bilinmeyen parametreleri göstermektedir. Eğer eksiklik,  $Y$ 'nin değerlerine bağlı değilse,

$$f(M | Y, \phi) = f(M | \phi) \quad (\text{tüm } Y, \phi \text{ 'ler için}) \quad (2.1)$$

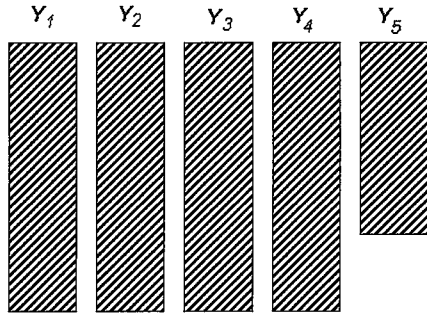
şeklinde gösterilir ve veriler TROK (MCAR) olarak adlandırılır. Burada dikkat edilmesi gereken varsayımın yapının kendisinin rassal olduğu anlamına gelmemesidir. Fakat, eksik olma durumu verilerin değerlerine bağlı değildir (24). Eksik veri, veri setinde bulunan herhangi gözlenen veya gözlenmeyen değerlerle ilişkili değildir (16).

### **2.3.2- Rassal Olarak Kayıp (ROK, MAR) Mekanizması**

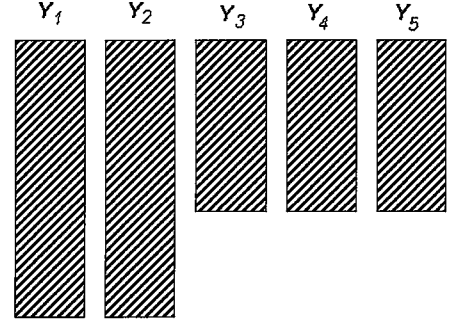
$A$  ve  $B$  gibi iki değişken ele alındığında, cevap olasılığı  $A$ 'ya bağlı fakat  $B$ 'ye bağlı değil ise, eksik veri ROK olarak düşünülür (24, 39).  $Y_{göz}$ ,  $Y$  veri setinin gözlenen kısmını,  $Y_{eks}$ 'de eksik olan kısmını göstermek üzere,

$$f(M | Y, \phi) = f(M | Y_{göz}, \phi) \quad (\text{tüm } Y_{eks}, \phi \text{ için}) \quad (2.2)$$

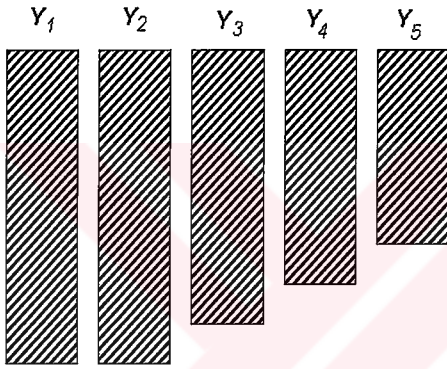
olacaktır. Bu varsayım TROK'dan daha az sınırlayıcı olan bir varsayımdır ve eksik olma sadece  $Y_{göz}$ 'e bağlıdır. Bu mekanizma ROK (MAR) olarak adlandırılır (15). ROK'da eksik olma durumu, eksik veri ile ilişkili değildir, fakat diğer gözlenen değişkenlerle ilişkilidir (16).



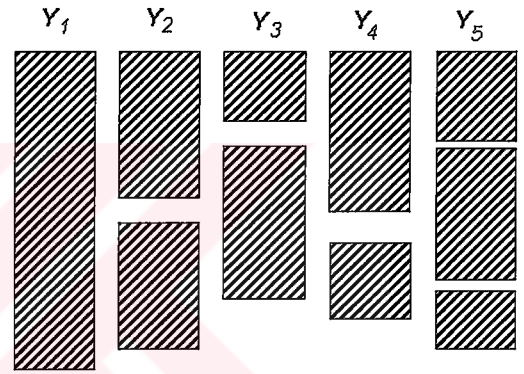
(a) Tekdeğişkenli cevapsız



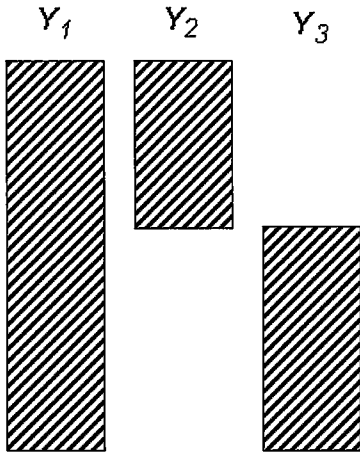
(b) Çokdeğişkenli iki pattern



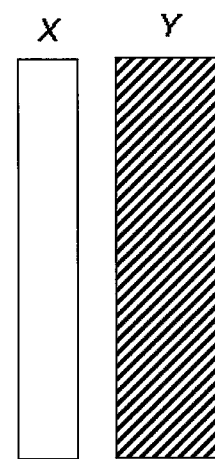
(c) Monoton



(d) Genel



(e) Dosya eşleştirme



(f) Faktör analizi

Şekil 2.1 Eksik veri yapıları

M'nin dağılımı, Y veri matrisindeki eksik değerlere bağlı ise bu mekanizma Rassal Olarak Kayıp Değil (ROKD), (not missing at random, NMAR) olarak adlandırılır. ROKD'de eksik olma durumu gözlenmemiş (missing) değere bağlıdır (16, 24).

Burada, belki de en basit veri yapısı, bazı birimleri eksik olan tekdeğişkenli rassal örneklemdir.  $Y=(y_1, \dots, y_n)^T$  tanımlansın; burada  $y_i$ , i. birim için rassal değişkenin aldığı değeri göstermektedir. Ayrıca  $M=(M_1, \dots, M_n)$  olarak tanımlanabilir; burada da  $M_i=0$  olması durumu, i. birime ait verinin gözlemlendiğini,  $M_i=1$  olması durumu ise, i. birime ait verinin eksik olduğunu gösterir.  $(y_i, M_i)$  ortak dağılımının birimlerden birbirinden bağımsız olduğu varsayılabilir, yani bir birimin gözlenmesi olasılığı diğer birimlerin Y veya M değerlerine bağlı değildir. Buradan yola çıkılarak

$$f(Y, M | 0, \phi) = f(Y | Y\phi) f(M | Y, \phi) = \prod_{i=1}^n f(y_i | \phi) \prod_{i=1}^n f(M_i | y_i, \phi) \quad (2.3)$$

olduğunu gösterilebilir. Burada  $f(y_i | \phi)$ ,  $y_i$ 'nin bilinmeyen parametreleri  $\phi$  ile indekslenmiş olasılık yoğunluk fonksiyonunu göstermektedir ve  $f(M_i | y_i, \phi)$   $y_i$ 'nin eksik olması olasılığının  $\Pr(M_i = 1 | y_i, \phi)$  olduğu ikili  $M_i$  göstergesi için Bernoulli dağılımının olasılık yoğunluk fonksiyonudur. Eğer eksik olma Y'den bağımsız ise yani  $\Pr(M_i = 1 | y_i, \phi) = \phi$   $y_i$ 'den bağımsız olan bir katsayı ise, eksik veri mekanizması MCAR olarak nitelenir (özellikle bu özel durumda ROK'a da eşit olur). Eğer mekanizma  $y_i$ 'ye bağımlı ise bu mekanizma ROKD'dir. Çünkü bu durumda ROKD eksik olan  $y_i$ 'lere bağımlıdır ve bu değerlerin az sayıda oldukları varsayılır (24).

$n_E$  cevap veren birimlerin sayısını gösterebilir ( $M_i=0$ ). Bu örnekteki eksik değer durumunda analizde birim sayısı, n örnekleme büyüklüğünden  $n_E$  örnekleme büyüklüğüne düşer. Bu durumda n birimlik örnekleme için yapılması düşünülen analizler,  $n_E$  birimlik örnekleme için uygulanır. Örneğin; verilerin normal dağıldığı varsayımı altında ortalama hakkında çıkarılabilir, ortalamasının tahmini,  $\frac{s}{\sqrt{n_E}}$  standart hatalı örnek ortalaması ile yapılabilir. Burada s, cevap veren birimlerin örnekleme standart sapmasını göstermektedir. Bu strateji; mekanizma, TROK mekanizması ise geçerlidir. Çünkü bu

durumda gözlenen değerler araştırma kapsamındaki bütün birimlerin rassal bir alt örneğidir. Fakat, veriler ROKD ise alt örneğe bağlı analiz genelde y dağılımının parametreleri için yanlıdır (24).

Çok sayıda eksik veri atama yöntemi bu varsayıma uygun olarak geliştirilmiştir. EM, regresyon ve çoklu atama (Multiple Imputation) yöntemi ROK durumu için uygun atama yöntemleridir (22).

### **2.3.3- İhmal Edilemez Kayıp (IEK, NI) Mekanizması**

A ve B gibi iki değişken ele alındığında, cevap olasılığının A ve B'ye bağlı olması mümkün ise, eksik veri İhmal Edilemez Kayıp (IEK, Nonignorable) mekanizması özelliğinde kabul edilir. Diğer bir ifadeyle 'eksik olma' rassal değildir ve veri setindeki bir diğer değişkenden tahmin edilemez. Atama yöntemlerinin tümü, IEK durumunda yanlı sonuçlara neden olurlar (15, 22, 39).

Veri setlerinde, eksik verinin rasgele koşullar altında olup olmadığını belirledikten sonra, giderilmesine yönelik yaklaşımlar incelenir. Eksik veri probleminin giderilmesinde kullanılan çözümler, eksik veri sürecinin rasgeleliği ve eksik veriyi tahmin etmede kullanılan yöntemlere dayalı olarak incelenir.

Veri setlerindeki eksik değerlerin TROK yapıda olduğu durumlarda uygulanabilecek yöntemler ile ROK yapıda olduğunda uygulanabilecek yöntemler farklıdır.

### **2.4- Eksik Verilerde Atama Yöntemleri**

Araştırmalardan elde edilen veri setlerinde, eksik verilerin bulunduğu araştırmalarda çözüm ve atama yöntemlerinin kullanılması gerekmektedir. Bu atama yöntemlerinin yanında liste veya durum bazında veri silme (listwise or casewise data deletion-LD or CD) ve çiftler bazında veri silme (pairwise data deletion-PD) gibi atama yapmadan eksik veri problemini ortadan kaldıracak çözüm yöntemleri de

kullanılmaktadır. Eksik veri yerine, ortalamanın atanması (mean substitution), regresyon ataması (regression imputation), hot deck, cold deck ataması (hot deck, cold deck imputation), EM yaklaşımı (expectation maximization (EM) approach) ve çoklu atama (multiple imputation) en çok kullanılan atama yöntemleridir (17).

Bu araştırmada sırasıyla, atama yapmadan eksik veri problemini çözmeye yönelik liste bazında veri silme veya durum bazında veri silme ile çiftler bazında veri silme ve yerine ortalamayı koyma yöntemlerine ve atama yöntemleri ele alınmış ve hangi durumlarda hangi yöntemlerin kullanılabileceğine yer verilmiştir. Ayrıca, bu değer atama yöntemlerinin kuvvetli ve zayıf yönlerine değinilmiştir.

#### **2.4.1- Tam Gözlemlerin Kullanılması Yöntemi (Listwise Method)**

Bu yöntemde, yalnızca eksik veri içermeyen (complete) birimlerin aldıkları değerler kullanılır. Diğer bir ifadeyle, veri setinde bir ya da daha fazla eksik veri olan gözlemler veriden çıkarılır. Bu yöntemin olumsuz tarafı, eksik veriye sahip gözlemlerin veriden çıkartılmasının neden olduğu bilgi kaybı olacaktır. İstatistik paket programlarının çoğunda yer alan ve eksik veri sorununun çözümünde en basit ve doğrudan bir yaklaşım olan bu yöntemin kullanılabilmesi için, eksik verinin TROK (MCAR) olması gerekmektedir (7).

Bu çözüm yöntemine göre, bir analizde kullanılan herhangi bir değişken eksik veri içeriyorsa, ilgili birim analizden tamamı ile çıkartılır. Başka bir anlatımla, tüm kayıtları olan yanıtlayıcı analizde yer alır ve diğerleri analiz dışı bırakılır. Bu yaklaşım eksik veriler için kullanılan en temel ve popüler yöntemlerden biridir, SAS ve SPSS gibi sıkça kullanılan istatistiksel paket programlarında yer almaktadır (2, 7, 13, 27). Çoğu istatistik paket programında eksik değerlere sahip birimlerin tamamı ile analizden çıkartılması seçeneği kurulu düzen (başlangıç önerisi, default option) olarak yer almaktadır (7).

Liste bazında silme yönteminde (Listwise Method) aşağıdaki formüllerden yararlanılmaktadır. Formüllerde j ve k indisleri kantitatif değişkenleri göstermektedir.

Liste bazında silme yöntemi ile ortalama;

$$\bar{x}^L = [\bar{x}_j^L] = \left[ \sum_i x_{ij} / n_c; \quad i \in I(J) \right] \quad (2.4)$$

formülüyle hesaplanır. Kovaryans ( $C^L$ ) ve korelasyon ( $R^L$ ) değerleri ise;

$$C^L = [c_{jk}^L] = \left[ \sum_i (x_{ij} - \bar{x}_j^L) * (x_{ik} - \bar{x}_k^L) / (n_c - 1); \quad i \in I(J) \right] \quad (2.5)$$

$$R^L = [r_{jk}^L] = [c_{jk}^L / (c_{jj}^L * c_{kk}^L)^{1/2}] \quad (2.6)$$

formülleri yardımıyla hesaplanmaktadır (13, 22).

#### **2.4.2- Eldeki Tüm Bilginin Kullanılması Yöntemi (Pairwise Method)**

Araştırmalarda elde edilen tam gözlemlerin kullanılması yöntemi, ortalamaların ve basit frekans dağılımlarının elde edilmesi gibi tekdeğişkenli analizler için olur olmaz başvurulan bir yaklaşımdır. Çünkü, bir değişkenin tam değeri, diğer değişken ya da değişkenlerde eksik veri olduğu için veriden çıkartılmaktadır (Listwise) (2).

Bu nedenle, eksik veri içeren araştırmalarda eldeki tüm bilginin kullanılması yöntemine sıklıkla başvurulur. Bu atama yöntemi, gerçekte eksik verilerin yerine konulması ya da eksik verilerin atanmış değerlerle doldurulması değil, eldeki tüm değerler yardımıyla dağılımı tanımlayıcı ölçülerin (ortalama, standart sapma vb) ve ilişki ölçülerinin (korelasyon, kovaryans) elde edilmesidir. İstatistik yazılımların çoğunda bu yönteme ilişkin menüler olduğu için araştırmacılar tarafından sıklıkla kullanılmış ve bu yaklaşımla elde edilen korelasyon ya da kovaryans matrisleri (faktör analizi) girdi matrisi olarak kullanılmıştır (2).

Bu yöntem, eldekilerin tümü yaklaşımı da (all available approach) denir. Bu yöntem, istatistiksel yazılımlarda (örneğin SPSS'de) "pairwise" seçeneği ile yapılır ve yukarıda belirtildiği gibi sıklıkla iki değişken arasındaki gözlem sayısını en büyükleterek (maksimize ederek) korelasyon katsayılarını elde etmek ve dolayısıyla örneklemdaki tüm ikili bilgiyi en büyükleme için kullanılır (2).



Pairwise atama yöntemi, toplam örnek hacmine bağlı olarak her bir değişkendeki eksik birimlerin sayısı az ve örnekte çok fazla değişken varsa atama yöntemleri içerisinde kullanılması en cazip yöntemlerden biridir. Bu yöntemde aşama aşama her bir değişkenin ortalama ve varyansları ayrı ayrı hesaplanır, ayrıca çift olarak ele alındıklarında ise ulaşılabilen tüm çiftlerin kovaryansları hesaplanır (7).

Tahminler, değişkenlerin her bir çifti için ayrı ayrı hesaplanır. Bu hesaplamalarda tüm birimler kullanılır (13).

Çiftler bazında silme (Pairwise Method) yöntemine ilişkin formül ve gösterimler aşağıdaki gibidir. Formüllerde j ve k indisleri kantitatif değişkenleri ve l indisi ise tüm değişkenleri göstermektedir.

Bu yöntemde ortalama;

$$\bar{x}^P = [\bar{x}_{ik}^P] = \left[ \sum_i x_{ik} / n_{ik}; \quad i \in I(1, k) \right] \quad (2.7)$$

formülüyle hesaplanır. Standart sapma ise;

$$\hat{\sigma}^P = [\hat{\sigma}_{ik}^P] = \left[ \left( \sum_i (x_{ik} - \bar{x}_{ik}^P)^2 / (n_{ik} - 1) \right)^{1/2}; \quad i \in I(1, k) \right] \quad (2.8)$$

formülünden yararlanılarak hesaplanmaktadır. Kovaryans ( $C^P$ ) ve korelasyonlara ( $R^P$ ) ilişkin formüller ise aşağıdaki gibidir.

$$C^P = [c_{jk}^P] = \left[ \sum_i (x_{ik} - \bar{x}_{jk}^P) * (x_{ij} - \bar{x}_{kj}^P) / (n_{jk} - 1); \quad i \in I(j, k) \right] \quad (2.9)$$

$$R^P = [r_{jk}^P] = [c_{jk}^P / (\hat{\sigma}_{jk}^P * \hat{\sigma}_{kj}^P)] \quad (2.10)$$

Bu yöntemin ayırt edici özelliği; her bir korelasyon için ayrı gözlem çiftlerinin ve her hesaplamada farklı sayıda gözlemin kullanılmasıdır. Bu yöntem yardımıyla elde edilen korelasyonlar bütün örneklemin temsilcisi olarak kabul edilir ve değer ataması yapılmaz. Tam gözlemlerin kullanılması yaklaşımında ise sadece tam (eksik verisi olmayan) gözlemler dikkate alınmaktadır. Eğer eksik veri süreci TROK (MCAR) değil

ise tam gözlemlerin kullanılması yönteminde olduğu gibi eldekilerin tümü yaklaşımında da elde edilen sonuçlar yanlı olabilmektedir (2).

Eldeki tüm bilginin kullanılması yöntemi, bir değişkendeki bir eksik nedeniyle o gözlemin tamamen veriden çıkarılmasına neden olan sorunu ortadan kaldırması ve kullanılan veriyi en fazlaya çıkarmasına rağmen bazı sorunları da beraberinde getirebilmektedir. Bu sorunların en belirginini, korelasyonların beklenen sınırların dışında elde edilmesi ve hesaplanan korelasyonların, korelasyon matrisindeki diğer korelasyonlar ile tutarsız olmasıdır. x ve y gibi iki değişken arasındaki korelasyon, bu değişkenlerin z gibi bir değişken ile yaptığı korelasyonlar ile sınırlı olup, bu sınırlar aşağıdaki gibi tanımlanır (2).

$$r_{xy}'\text{nin sınırları} = r_{xz} r_{yz} \pm \sqrt{(1-r_{xz}^2)(1-r_{yz}^2)} \quad (2.11)$$

Matristeki korelasyonlara ilişkin sınırları tanımlayan bu formüle göre, x ve y'nin diğer değişkenlerle aralarındaki korelasyonlar arttıkça, x ile y arasındaki korelasyonun olası sınırları azalır. Bu durum, bir gözlem kümesinden elde edilen korelasyonun diğer gözlem kümesinden elde edilen korelasyonlarla tutarsız olması olasılığını artırır. Örneğin; x ile y'nin, z gibi bir değişkenle korelasyonu sırasıyla 0.70 ve 0.30 olsun. Buna göre x ile y arasındaki korelasyonun olası sınırları;

$$r_{xy}'\text{nin sınırları} = 0.70 \times 0.30 \pm \sqrt{(1-0.70^2)(1-0.30^2)} = [-0.47; +0.89] \quad (2.12)$$

olarak bulunacaktır. Dolayısıyla, bu sınırlar dışında bir korelasyon katsayısının elde edilmesi, sonuçların matematiksel olarak tutarsızlığını gündeme getirmektedir. Bu durumun, iki korelasyonun gözlem kümelerinin ve sayılarının değişiklik göstermesi durumunda ortaya çıkması oldukça olası olup, bu durum eldekilerin tümü yaklaşımının bir sonucudur (2).

Korelasyon katsayılarındaki bu soruna bağlı olarak ortaya çıkan bir diğer sorun ise, korelasyon matrisinin özdeğerlerinin (eigenvalues) sıfır'dan küçük (negatif) çıkabilmesidir. Bu durumda, korelasyon matrisinin varyans özellikleri değişir. Bu nedenle, korelasyon matrisleri düzeltilebilir. Bazı istatistik yazılımlar düzeltilmiş korelasyon matrislerini de çıktı olarak verebilmektedir. Bu sorunların tümü, eksik

veriye eldekilerin tümü yaklaşımı ile çözüm arandığında dikkate alınmalıdır. Bilindiği gibi tüm özdeğerleri pozitif olan matrisi Pozitif Tanımlı Matris (positive definite matrix) denir. Bir matrisin pozitif tanımlı olmamasının nedenleri arasında; veride oldukça fazla eksik gözlemin olması ve korelasyon/ kovaryans matrislerinin eldekilerin tümü yaklaşımıyla elde edilmesi, değişkenler arasında doğrusal bağımlılığın olması, herhangi bir değişken değerlerinin sabit olması, gözlem sayısının az olması gibi nedenler söz konusudur. Ayrıca bu yöntemde karşılıklı olarak birbirini tutmayan, çelişkili korelasyonlar ile korelasyon matrislerinin hesaplanması mümkündür (7).

İstatistiksel yazılımların çoğu, eksik veri analizi menüsünün Pairwise yöntemi seçeneğinde; ikişerli gözlem sayıları matrisi, kovaryans matrisi ve korelasyon matrisi ile eksik gözlemler veriden çıkartıldığında diğer değişkenlerin ortalama ve standart sapmalarında ne miktarda bir değişim olduğunu gösteren tabloları çıktı olarak vermektedir (13).

#### **2.4.3- Gözlemlerin ve/veya Değişkenlerin Silinmesi Yöntemi**

Veri setlerindeki eksik verilerin giderilmesine yönelik bu yaklaşım, eksik veri içeren gözlem(ler)in ve/veya değişken(ler)in tamamen veri setinden silinmesidir. Fakat, veriden çıkarılacak uygun eksik veri ya da değişken miktarının ne olacağı konusunda kesin bir yöntem bulunmamaktadır. Bu yöntem, sıklıkla, eksik verilerin küçük bir gözlem grubunda ve/veya değişkende yoğunlaşması ve bunların veriden çıkartılmasının eksik veriyi büyük oranda azaltması söz konusu olduğunda kullanılır. Araştırmacı, gözlem sayısının fazla olduğu durumlarda, gözlem ve/veya değişkenleri veriden çıkartma yoluna gidip gitmeyeceğine iyi karar vermek zorundadır. Fakat, araştırmacının veri setinde herhangi bir değişkeni kaybetmek istememesi de çok normal bir durumdur.

Bazen, özellikle eksik verinin az sayıda olduğu ve/veya az sayıda değişkende yoğunlaştığı ve eksik veri yapısının rasgele olmadığı durumlarda da bu yöntemle başvurularak rasgele bir yapıya ulaşılmaya çalışılabilir. Eksik verilerin, "yapılacak istatistiksel çözümlemede bağımlı değişken olan değişkende" olması durumunda, bu

gözlemlerin veriden çıkartılması önerilmektedir. Böylece, değer atama yöntemlerinden (imputation methods) biri ile eksik veriler yerine değer ataması yapıldıktan sonra yapılacak olan çözümlemede, değer atanmasının yol açacağı yapay açıklayıcılık artışının önüne geçilebilir.

Eğer eksik verilerin oldukça fazla olduğu bağımsız bir değişken söz konusu ve araştırmacı bu değişkeni veriden çıkarma eğiliminde ise bu değişkenin diğer bağımsız değişkenlerle yüksek derecede ilişkili olmasına özen gösterilmelidir. Genel bir yaklaşım olarak, çokdeğişkenli çözümlemede, bir değişkenin silinmesine karşı eksik veri kaynağının ortadan kaldırılmasının getireceği kazanç her zaman göz önünde bulundurulmalıdır (2, 22, 38, 39).

## **2.5- Değer Atama Yöntemleri (Imputation Methods)**

Eksik verinin giderilmesine yönelik olarak başvurulacak bir yöntem; örneklemdaki tam veri içeren değişken ve/veya gözlemlerin değerlerine dayanarak eksik verilerin tahmin edilmesidir. İstatistiksel çözümler, eksik verileri tahmini değerlerle doldurulan (tamamlanan) yeni veri seti yardımıyla yapılır. Veri seti içerisinde yer alan eksik verilerin kestirilmesi için geliştirilmiş birçok yöntem vardır. Bu yöntemler değer atama yöntemleri olarak tanımlanırlar (2, 13, 22, 36, 38, 46).

Değer atama yöntemleri çekici olduğu kadar sonuca etkileri nedeniyle de tehlikeli olabilecek yöntemlerdir. Değer atama yöntemleri; örneklemin tümüne genellenmek amacıyla gözlemlerin bir alt kümesinden elde edilen tüm bilginin kullanılması ya da eksik veri yerine atanacak değerlerin kestirilmesi şeklinde tanımlanabilirler. Değer atama yöntemlerinden bazıları,

- Eksik Gözlem ile Tam Gözlemin Yer Değiştirmesi (Case Substitution)
- Eksik Gözlem ile Ortalamanın Yer Değiştirmesi (Mean Substitution)
- Hot Deck Atama (Hot Deck Imputation)
- Cold Deck Atama (Cold Deck Imputation)
- Regresyon Atama (Regression Imputation)
- EM Atama (Expectation Maximization)
- Çoklu Atama (Multiple Imputation)
- DA Atama (Data Augmentation) olarak sayılabilir.

### **2.5.1- Eksik Gözlem ile Tam Gözlemin Yer Değiřtirmesi (Case Substitution)**

Bu yöntemin en yaygın örneđi, örnekleme giren (örneklem çerçevesinde yer alan) ancak kendisine ulařılmayan (verileri tamamen eksik olan) ya da bilgilerinde çok fazla eksik olan kiřinin, örneklemede olmayan ancak "örneđe giren ve kendisine ulařılmayan kiřinin özelliklerine benzer" bir kiři ile yer deđiřtirmesidir. Kısaca, yedek denek atama řeklinde de özetlenebilir (2, 22).

### **2.5.2- Eksik Gözlem ile Ortalamannın Yer Deđiřtirmesi (Mean Substitution)**

Eksik veri yerine atama yapmak için sıklıkla kullanılan yöntemlerden biri, incelenen deđiřkene iliřkin eksik verilerin yerine, o deđiřkene iliřkin tam gözlemlerin (complete) ortalamasının atanmasıdır (7, 8, 22, 27, 30). Bu iřlem sonucunda, tam ve atanmış veriler için elde edilen ortalama, tam verilerle elde edilen ortalamaya eřit çıkacaktır. Bu yöntemin olumlu yönü, kolay uygulanabilir olması ve tam bilgiye sahip gözlemleri dikkate almasıdır.

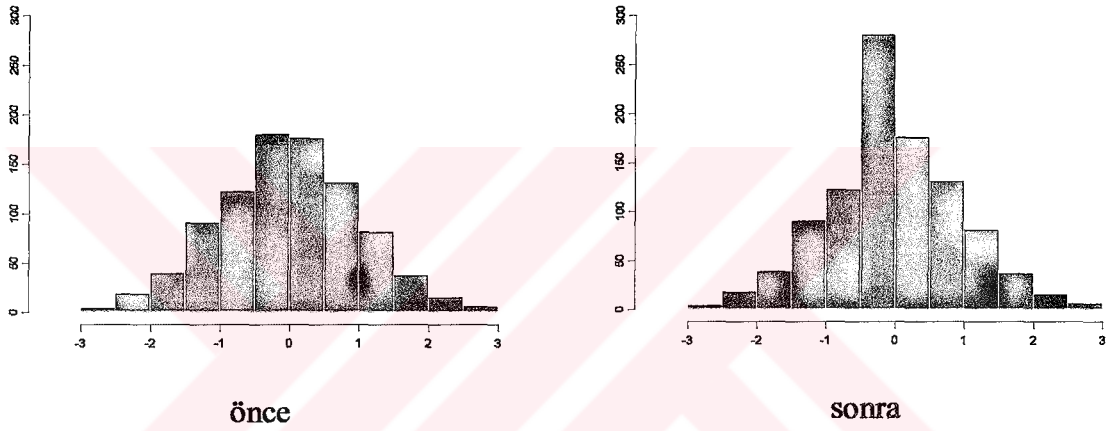
Bu yöntemin olumsuz yönlerini Little ve Rubin 1987 yılında belirtmişlerdir. Bunlar;

1. Örnek hacminin abartılı bir řekilde büyümesi,
2. Varyansların olduđundan daha küçük çıkması,
3. Kovaryansların negatif eğilimli olması ve,
4. Atama sonrasında elde edilen yeni deđiřkenin dađılımının toplumu yeterince yansıtamamasıdır. Çünkü asıl deđiřkene ait dađılım, ortalamaya eřit olan deđerlerin eklenmesiyle deđiřmektedir (7, 22).

Örneđin, deđiřkenlerdeki tüm eksik gözlemlerin yerine tek bir deđer (ilgili deđiřkenin ortalaması ile) atandıđı için deđiřkenler gerçek dađılımından uzaklařır ve verideki gerçek varyans olduđundan küçük çıkar (Tam ve atanmış verilerin varyansı  $[(n^{(j)} - 1)/(n - 1)]s_{jj}^{(j)}$  ile verilir. Burada n: toplam gözlem sayısı,  $n^{(j)}$ : tam gözlemlerin sayısı,  $s_{jj}^{(j)}$  ise tam gözlemlerin varyansıdır. Böylece, tam ve atanmış deđerlerle

tamamlanmış veri setlerine ilişkin dağılımın varyansı  $(n^{(j)} - 1)/(n - 1)$  çarpanı derecesinde olduğundan eksik bulunur.) Ayrıca, varyanstaki küçülmeye bağlı olarak bu değişken ile diğer değişkenler arasındaki korelasyon katsayıları daha düşük elde edilir (2, 22).

Bu durum bazı istenmeyen problemlere neden olabilir. Değişkenin dağılımına ait dağılım grafiği bu işlemten önce ve sonra aşağıdaki gibi olur. Eksik değerlerin yerine ortalama atandığında değişkene ait dağılım ve  $X$ 'in marjinal dağılımı bozulur,



Şekil 2.2: **Önce:** Eksik veri yapısına ait dağılım grafiği, **Sonra:** Eksik gözlem değerlerinin yerine ortalamanın atıldığı durumda veri yapısına ait dağılım grafiği.

Bununla beraber,  $X$  değişkeni ile diğer değişkenler arasındaki kovaryans ve korelasyonlar da bozulur (3, 22, 40).

### **2.5.3- Hot Deck Atama (Hot Deck Imputation) Yöntemi**

Hot deck ataması cevapsız değişkenler için sıklıkla kullanılan bir yöntemdir ve bazı avantajları vardır: Bu avantajlardan en önemlisi değişkenlere ait dağılımı muhafaza etmesidir. Bütün değişkenler için aynı örnek tartısını kullanır ve farklı analizlerden elde edilmiş sonuçlar bir diğeriyle tutarlıdır (41).

Hot deck yönteminin ana prensibi, eksik veriler için elde mevcut olan veriyi kullanarak uygun atamayı yapmaktır. Veri seti içerisinde bulunan, eksik veriye sahip

birimin yerine, o birime benzer (donör) birimin konması anlamına gelir. Bu eşleştirme işlemi, bir filtre değişkeni kullanılarak yapılır. Veri setindeki tüm kayıtlar filtre değişkenindeki kayıtlarla karşılaştırılır. Eksik cevaba atama yapmak için eldeki mevcut veri setinden atama yapılır. Diğer hot deck atama yöntemleri, uzaklık eşleme fonksiyonu veya en yakın komşu atama fonksiyonunu kullanır (41).

Hot deck atama yönteminde, veri matrisindeki eksik gözlemler benzer gözlemlerle doldurulur.

Hot deck atama yöntemi uzun zamandan beri kullanılmaktadır. Bu atama yöntemi, liste bazında veri silme, çiftler bazında veri silme, yerine ortalamayı koyma yöntemlerinden daha üstün bir yöntemdir. Hot deck atama yönteminin avantajları arasında kavramsal basitliği, değişkenlerin ölçüm düzeylerini koruması (kategorik değişkenler kategorik olarak, sürekli değişkenler sürekli olarak kalır) ve tamamlanmış veri matrisi elde edilmesi sayılabilir.

Tamamlanmış veri matrisi sayesinde de standart istatistiksel analizler uygulanabilir.

Hot deck atama yönteminin en önemli dezavantajı, 'benzerlik' kavramının tanımlanmasındaki güçlüktür. Bu nedenle hot deck prosedürü eksik veriler için standart bir yol sağlamamaktadır. Bu benzerliğin belirlenebilmesi için, uygun birimlerin seçimini başarabilecek bir yazılım gerekmektedir. Daha ileri bir hot deck algoritmasına göre, benzer bir kayıttan daha fazla sayıda kayıt belirlenir ve uygun birimlere ait kayıtlardan biri, eksik olan değerlerin ataması için rassal olarak seçilir. Ayrıca eğer uygunsa, bu uygun birimlerin ortalaması eksik verilerin atanması için kullanılır (27).

#### **2.5.4- Cold Deck Atama (Cold Deck Imputation) Yöntemi**

Bu yöntemde; önceki çalışmalardan, aynı çalışmanın bir önceki uygulamasından ya da dış kaynaklardan elde edilen sabit bir değer, ilgili değişkendeki tüm eksik veriler yerine atanır. Bu yöntemin eksik gözlem ile ortalamanın yer değiştirmesi yönteminden

tek farkı, atanacak değerin kaynağıdır. Araştırmacı, dış kaynaklardan elde edilen değerin, eldeki veriler yardımıyla elde edilen ortalama gibi bir değerden daha geçerli olabileceğinden emin olmalıdır. Bu yöntem, ortalama atama yöntemine benzer olumsuzluklara sahiptir (2).

### **2.5.5- Regresyon Atama (Regression Imputation) Yöntemi**

Çoklu regresyon atama yöntemi, eksik verilere atama yapmak için kullanılan diğer bir yöntemdir. Bu yöntem, eksik veri içeren her bir değişkenin, diğer tüm değişkenler üzerine veya ilişkili değişkenler üzerine regresyon denkleminin kurulmasını içerir (48).

Bu atama yöntemi oldukça karmaşıktır. Çünkü; değişkenler arasındaki ilişkiyi göz önüne almaktadır. Regresyon atama yöntemi eksik veri içermeyen değişkenlerdeki bilgiyi kullanarak eksik veri içeren değişkenlerin tahminlerini ya da atamalarını yapar. Eksik veri içeren her bir değişken, kriter değişkeni olarak işlem görür ve diğer bütün değişkenler bağımsız değişken olarak ele alınarak, kriter değişkeni tahminlemek için regresyon analizine tabi tutulur (7).

Bağımsız değişken sayısına ve veri setindeki eksik verinin oranına bağlı olarak regresyon yönteminde bir çok değişik yaklaşımlar vardır. Doğrusal regresyon analizi yönteminin uygulanmasındaki varsayım, tamamlanmamış değişken ile bağımsız değişkenler arasındaki ilişkiye bağlı olarak yapılmaktadır (7).

Bu regresyon denkleminde elde edilecek tahmini değer, eksik verilere atamada bulunmak için kullanılır. Regresyon atama yöntemi kullanıldığında bir rassal hatanın modele dahil edilmesi önemlidir. Bu hata terimi ya rassal olarak belirlenir, ya bir rassal normal sapmadır, ya da kullanıcı tarafından belirlenecek serbestlik derecesine bağlı bir rassal t değeri olabilir. Bu atama yöntemi, bağımsız değişkenlerdeki eksik verilerin atanması için kullanıldığında, bu durum çoklu doğrusal bağımlılığa katkıda bulunacaktır. Çünkü eksik veriler için atanan değerler, modeldeki diğer değişkenler ile ilişkili olacaktır. Regresyona dayalı atama yönteminde, modele dahil edilmeyen diğer



değişkenlerin kullanılması da mümkündür. Fakat, bu durumda daha zayıf tahmin değerleri elde edilecektir (27, 48).

Regresyon atama yöntemini kullanmanın bir avantajı, yerine atama yapılacak eksik veri içeren değişkenin, her bir eksik verisi için farklı bağımsız değişkenler kümesini kullanmasıdır. Bu yaklaşımın, yerine ortalamayı koyma yaklaşımından bir üstünlüğü, eksik veri içeren değişkenlerin varyans ve kovaryanslarını korumasıdır. Çünkü bir değişkenin eksik her bir durumu, diğer değişkenlerin değerlerine bağlıdır ve her seferinde farklı bir tahmin değerini verecektir. Bu yöntemin EM yaklaşımından dezavantajı ise daha sınırlı sayıda bilgiyi kullanmasıdır (27).

Regresyon atama yönteminin, eksik verinin orta (ılımlı) düzeyde olduğu ve veride yaygın bir dağılım gösterdiği durumlarda kullanılması önerilmektedir (2). Ayrıca, değişkenler arasındaki ilişki yeterli olmalı ve iyi bir kestirim yapabilmek için bağımlı değişken, bağımsız değişken(ler) tarafından yeterince açıklanmalıdır. Böyle bir açıklanma söz konusu değil ise eksik gözlemlere ortalamanın atanması daha uygun olacaktır. Bu yöntemin olumsuz tarafları da vardır;

- 1) Bu yöntemde değişkenler arasındaki ilişkiden yola çıkıldığı için atama sonucunda veride zaten var olan ilişki daha da kuvvetlenmiş olur,
- 2) Bu yöntemin kullanımı arttıkça, elde edilen veri kendine özgü bir yapıya ulaşabilir ve genellenebilirliği azalır,
- 3) Regresyon yöntemi sonucunda elde edilen kestirimin sınırları, veride var olan sınırların ötesinde bulunabilecektir (örneğin, bağımlı değişken değerleri 0 ile 20 arasında değişiyor ise bu aralıktaki bir değer için kestirim değeri 21 olarak bulunabilir). Bu nedenle bazı düzeltme işlemlerinin yapılması gerekli olabilir (2). Regresyon yönteminde eksik veriler çoklu doğrusal regresyon kullanılarak tahmin edilebilir. Tahmini değişkenlerin ortalamaları, kovaryans matrisi ve korelasyon matrisi hesaplanabilir. Regresyon yönteminde, regresyon tahminlerine rasgele bir bileşen (component) eklenebilir. Artıklar, normal olasılıklı değişken, student t olasılıklı değişkeni veya düzeltme seçeneklerinden biri seçilebilir (13).

Regresyon atama yönteminde eksik verilerin tahmininde aşağıdaki formül ve gösterimlerden yararlanılmaktadır. Bu formüllerde j ve k indisleri kantitatif değişkenler için, l indisi ise tahmini değişkenler için kullanılmıştır.

$$x_{ij}^R = \begin{cases} x_{ij} & \text{eğer } x_{ij} \text{ kayıp değilse} \\ \text{regresyon tahmini } x_{ij} & \text{eğer } x_{ij} \text{ kayıpsa} \end{cases}$$

Regresyon tahmini  $x_{ij}$

$$x_{ij}^R = \beta_{0,ij} + \sum_1 \beta_{l,ij} * x_{il} + \varepsilon_{ij} \quad l \in J_1 = J(1: x_{il} \text{ kayıp değilse ve } l \neq j) \quad (2.13)$$

formülüyle hesaplanır. Burada;

$[\beta_{0,ij}, \beta_{l,ij}]$ ;  $\text{Diag}(\bar{X}^P) = [\bar{x}_{ij}^P]$ 'dan,  $C^P$ 'nin  $J_1$  köşegenlerinin en iyi (best) q dönüşümü yapılarak hesaplanır. Buradaki en iyi, ileriye doğru aşamalı seçimdir (forward stepwise selection). q ise maksimum bağımsız değişken sayısına eşit ya da küçük bir değere sahiptir. Aynı zamanda q değeri, belli bir F değeri belirlenerek de sınırlandırılabilir.

$\varepsilon_{ij}$ , rassal hata terimidir ve aşağıdaki dört farklı şekilde tanımlanır.

- Rasgele seçilmiş tam (eksiksiz, complete) veri setinden elde edilmiş artık değerlerdir (residuals),
- Tahminin standart hatası ile ölçülmüş rassal normal dağılım gösterir,
- Tahminin standart hatası ve serbestlik derecesi ile ölçülmüş rasgele t(sd) dağılımı gösterir,
- Ya da hata terimi ayarlaması olmayan bir durumdur.

Her bir eksik  $x_{ij}$  değeri için regresyon katsayıları kümesi  $(\beta_{0,ij}, \beta_{l,ij})$  ve  $\varepsilon_{ij}$  hesaplanır. Burada ortalama;

$$\bar{x}^R = [\bar{x}_j^R] = \left[ \sum_i x_{ij}^R / n; \quad i \in I \right] \quad (2.14)$$

formülüyle hesaplanır. Değişkenlere ilişkin kovaryans;

$$C^R = [c_{jk}^R] = \left[ \sum_i (x_{ij}^R - \bar{x}_j^R) * (x_{ij}^R - \bar{x}_j^R) / (n-1); \quad i \in I \right] \quad (2.15)$$

formülü, korelasyonlar ise;

$$R^R = [r_{jk}^R] = [c_{jk}^R / (c_{jj}^R * c_{kk}^R)^{1/2}] \quad (2.16)$$

formülü yardımıyla hesaplanır.

### **2.5.6- EM (Expectation-Maximization) Yöntemi**

EM algoritması eksik verilere sahip birimlerin tahmininde kullanılan maksimum benzerlik tahminlerini içeren bir yöntemdir (11, 15, 17, 26). EM yöntemi birçok araştırmacı tarafından sıklıkla kullanılmaktadır. EM yöntemi, yinelemeli (iterative) ve iki aşamalı bir yöntem olup E aşaması eksik veri için en iyi olası kestirimleri, M aşaması ise eksik veri atandığında ortalama, standart sapma ya da korelasyona ilişkin kestirimleri verir. Bu süreç, kestirilen değerlerdeki değişimin önemsenmeyecek derecede azalmasına kadar devam eder (2, 15, 17).

Bu yaklaşımın kullanılması aşamasında, tahmin edilen değere eklenmek üzere rassal olarak bir hata değerinin belirlenmesi gerekmektedir. Eğer bir regresyon denkleminde elde edilen tahmin güçlü bir tahmin ise (örneğin  $R^2$ 'si büyükse), denkleme küçük bir hata değeri eklenir. Tersine bir regresyon denkleminde elde edilen tahmin güçlü bir tahmin değilse, büyük bir hata değeri denkleme eklenir. Bu hata değerlerini elde etmenin kolay bir yolu, örneğin  $X_1$  değişkeni eksik veri içermiyorsa,  $X_1$  değişkeni için elde edilecek artıkları (residual) bu hata değeri için kullanmaktır. Bu hata değerlerinden biri atanacak değere eklenmek veya çıkarılmak üzere rassal olarak seçilir. Böylece bu tekrarlı prosedürün ilk adımı tamamlanmış olur. Sonuçta eksik veri içermeyen bir veri matrisi elde edilmiş olur ve bu yeni veri matrisine dayalı olarak kovaryans matrisi hesaplanır.

Bu prosedürler doğrusallık ve normal dağılım koşuluyla uygulanır. Fakat özel tahmin yöntemleri, bilinen herhangi bir dağılıma uygun olan kategorik veriler veya kategorik olmayan veriler için kullanılabilir. EM algoritmasını kullanan yazılım

programları, daha çok normallik varsayımına dayanmakta ve atama için regresyonu kullanmaktadır (13, 26, 27).

Yukarıda sözü edilen diğer atama yöntemleri gibi, EM yaklaşımı da geçerli olmayan standart hatalara yol açar. EM yaklaşımı, çoklu atama yöntemi için genişletildiğinde, standart hataların geçerli tahminlerini elde edilmesi mümkün olabilir (22, 27, 35).

EM yönteminde tekrarlı işlemler (iterative process) kullanılarak eksik veriler içeren kantitatif değişkenlerin ortalamaları, kovaryans matrisi ve korelasyonu tahmin edilir (26). Verilerin dağılımı için; normal, karışık (mixed) normal ve student t dağılımı gibi bir çok değişik varsayım yapılabilir. Burada karışık normal varsayımı için bir oran (proportion) ve standart sapma oranı (Standard deviation ratio); Student t testi için ise serbestlik dereceleri (degrees of freedom) açıkça belirtilmelidir (13).

EM atama yönteminde kullanılan formül ve gösterimler aşağıdaki gibidir. Bu formüllerde j ve k indisleri kantitatif değişkenleri ve l indisi ise tahmini değişkenleri göstermektedir.

Eksik verilerin tahminleri, ortalama vektörü ve kovaryans matrisi;

$$x_0 = [\bar{x}_j^0] = \text{Diag}(\bar{x}^P) = [\bar{x}_{jj}^P] \quad (2.17)$$

$$C_0 = [c_{jk}^0] = C^P = [c_{jk}^P] \quad (2.18)$$

formülleriyle hesaplanır.

Burada m=1 den M'ye kadar veya yakınsama elde edilinceye kadar. Eğer  $x_{ij}$  gözlem değeri eksik değilse (tam ise)

$$x_{ij}^m = x_{ij} \quad (2.19)$$

eğer  $x_{ij}$  eksikse m. iterasyon sonucu tahmin edilen değer

$$x_{ij}^m = \beta_{0,ij}^{m-1} + \sum_1 \beta_{l,ij}^{m-1} * x_{il}; \quad 1 \in J_2 = J(1: x_{il} \text{ kayıp değilse ve } l \neq j) \quad (2.20)$$

formülü yardımıyla hesaplanır. Burada  $[\beta_{0,ij}^{m-1}, \beta_{1,ij}^{m-1}]$  değerleri  $\bar{x}_{m-1}$  ve  $C_{m-1}$ 'den hesaplanır.  $\bar{x}_m$  ve  $C_m$  aşağıdaki gibi hesaplanır.

$$\bar{x}_m = [\bar{x}_j^m] = \left[ \sum_i w_i * x_{ij}^m / \sum_i w_i; \quad i \in I \right] \quad (2.21)$$

$$C_m = [c_{jk}^m] = \left[ \frac{\sum_i w_i * x_{ij}^m (x_{ij}^m - \bar{x}_j^m) * (x_{ik}^m - \bar{x}_k^m) + \sum_i \sum_{s \in J_2} c_{j,s|J_2}^{m-1}}{(n-1) * \sum_i w_i / n}; \quad i \in J_2, s \notin J_2 \text{ ve } s \neq j \right] \quad (2.22)$$

burada  $c_{j,s|J_2}^{m-1}$ ,  $C_{m-1}$  değerini döndüren (pivot eden)  $J_2$  nin j. satır s. elemanıdır.

$$w_i = \begin{cases} 1 & \text{çokdeğişkenli normal için} \\ \frac{1 - \alpha + \alpha * \lambda^{1+p/2} * \exp((1-\lambda) * D^2 / 2)}{1 - \alpha + \alpha * \lambda^{p/2} * \exp((1-\lambda) * D^2 / 2)} & \text{bozulmuş normal için} \\ (sd + p) / (sd + D^2) & \text{t(sd) dağılımı için} \end{cases} \quad (2.23)$$

$\alpha$  = Bozulma oranı

$\lambda$  = Standart sapma oranı

$p$  = Bağımsız değişkenlerin (tahmincilerin) sayısı =  $J_2$  deki indislerin sayısı

$D^2$  = Birimin ortalamadan olan Mahalanobis uzaklığının karesi

$$D^2 = \sum_{jk} (x_{ij}^m - \bar{x}_j^m) * (c_{jk}^m)^{-1} * (x_{ik}^m - \bar{x}_k^m) \quad (2.24)$$

Yukarıdaki formülde  $(c_{jk}^m)^{-1}$  ifadesi  $C_m^{-1}$ 'in jk. elemanıdır.

Algoritmanın sona ermesi için belirli bir değer ile kontrol edilmesi gerekmektedir. Örneğin regresyon analizinde artık kareler toplamı arasındaki fark çok küçük bir değere eşit olursa iterasyona devam edilmez ve o andaki parametre tahminleri en iyi değerler olarak ele alınır. Analizimizde yakınsama için

$$\left| c_{ij}^m - c_{ij}^{m-1} \right| / c_{ij}^m \leq \text{Yakınsama} \quad (2.25)$$

koşulunun geçerli olması aranacaktır.

Veri atama (doldurma);

$$X_i^E = [X_{ij}^E] = [X_{ij}^{m'}] \quad (2.26)$$

burada  $m'$ ,  $m$ 'nin son değeridir.

Ortalama ve kovaryanslara ilişkin formüller aşağıdaki gibidir.

$$\bar{X}^E = [\bar{X}_j^E] = \bar{X}_{m'} = [\bar{X}_j^{m'}] \quad (2.27)$$

$$C^E = [c_{jk}^E] = C_{m'} = [c_{jk}^{m'}] \quad (2.28)$$

Korelasyon ise,

$$R^E = [r_{jk}^E] = [c_{jk}^E / (c_{jj}^E * c_{kk}^E)^{1/2}] \quad (2.29)$$

formülü yardımıyla hesaplanır (3, 13, 22).

### **2.5.7- Çoklu Atama (Multiple Imputation-MI)**

Bu atama yönteminde, eksik veriler, iki ya da daha fazla yöntemin birlikte kullanılması sonucunda tahmin edilir. Dolayısıyla, çoklu atama yöntemi, karma bir kestirim değeri elde etmeyi amaçlar. Bu karma kestirim değeri, genellikle, iki ya da daha fazla yöntemle elde edilmiş kestirim değerlerinin ortalamasıdır. Bu kestirim tek bir yöntemle elde edilen kestirim değerine göre daha güvenilir olacağı düşünülmektedir (2, 14).

Çoklu atama yöntemi (MI), eksik verilerin yerine,  $m$  tekrar sayısı ve  $m > 1$  olmak üzere simüle edilmiş versiyonlarının kullanıldığı bir Monte Carlo tekniğidir. Buradaki  $m$  sayısı oldukça küçüktür (3-10 arasında) (37). Bu yöntem üç temel adımı gerektirir;

- Atama (imputation),
- Analiz etme (analysis) ve,
- Bir araya getirme (pooling).

Bu adımlar arasında başarılması en zor olanı atama adımdır. Bu adımda karşılaşılabilecek problemler aşağıdaki gibidir (14, 37).

- Herhangi bir gözlemin eksik olması, o gözlemin değerine bağlıdır. Örneğin yüksek veya düşük gelir düzeyine sahip kişiler gelir sorusunu atlama eğilimindedir.
- Eksik veriler, veriler kümesinin her hangi bir yerinde görülebilir.
- Atama adımı kullanılarak kullanılan yöntem, daha sonra yapılması düşünülen tamamlanmış veriler ile analiz aşamasının öngörülmesini zorunlu kılar (27).

Atanan veriler için tekrar uygulanan analiz adımı, atama uygulanmadan önce yapılan aynı analizden daha basittir. Çünkü eksik veri problemi ortadan kalkmıştır. Bir araya getirme adımı ise, m defa tekrarlanmış analizlerden, p değerleri, güven aralıkları, varyanslar ve ortalamaların hesaplanmasını içerir. Bu hesaplamalarda, genel olarak basit hesaplamalardır.

MI'lar yaratacak yeni analizsel yöntemlerin ve yazılımların oluşumuyla, eksik verilerle engellenen biomedikal ve sosyal bilim araştırmacıları için artan bir biçimde çekici bir hal almaktadır (27).

Çoklu atama yönteminin çok sayıda avantajı bulunmaktadır. En önemlisi, anlaşılır bir yöntemdir. Aynı zamanda analizde yer alan değişkenlerin normalliğinin ihlal edildiği durumlarda da robust sonuçlar vermektedir. Liste bazında veri silme, çiftler bazında veri silme ve yerine ortalamayı koyma yöntemlerinden pek çok durumda üstün olmaktadır. Dezavantajı ise üçten ona kadar veri kümesinde atama işlemi yaparken yoğun zaman gerektirmesidir (27).

### **2.5.8- DA (Data Augmentation) Yöntemi**

DA algoritmaları MCMC (Monte Carlo Markov Chain) yöntemleridir. Bu yöntemler Monte Carlo yöntemlerine benzer bir yöntemdir. Bilinmeyen  $\pi(x)$  dağılımının parametreleri; ya bu dağılımdan örnekleme yapılarak ya da uygun bir şekilde seçilmiş bir dağılımdan yeniden ağırlıklandırılmış örneklemeler seçilerek tahmin edilir. Genel ve yüksek boyutlu dağılımlarda Monte Carlo yöntemlerinin uygulanması

neredeysse imkansız hale gelmektedir. MCMC yöntemleri örnekleme yapılmasını kolaylaştıracak bir durum uzayı oluşturulmasına olanak vererek MC yöntemlerinin uygulanabilirliğini arttırmaktadır.

Burada oluşturulan zincir uzun bir süre çalıştırılırsa, bu zincirin benzetim sonucu değerleri,  $\pi(x)$ 'in özelliklerinin açıklayıcı veri analizi yöntemleri kullanılarak özetlenmesine olanak verir. Zincirlerin oluşturulması için bir çok yöntem geliştirilmiştir. Bu algoritmalardan bazıları, Gibbs örnekleme, Tanner ve Wong (1987) DA yöntemleri, sequential atamadır. Gibbs örnekleme, bir çok yöntemin uygulanmasının çok zor olduğu durumlarda dahi kolaylıkla uygulanabilir. Gibbs örneklemesinin başarılı olmasının nedeni problemin daha basit problemler dizisi haline getirilmesidir. Her bir problem bir seferde, bir bilinmeyen özellik için çalıştırılır. Her bir bilinmeyen özellik, daha sonra tam koşullu dağılımdan örneklenir. Eksik veri problemlerinde hem  $\theta$  parametreleri, hem de eksik veri  $Y_{\text{Eksik}}$  bilinmemektedir. Çünkü,  $\theta$ 'nın ve  $Y_{\text{Eksik}}$ 'in, ortak sonsal (posterior) dağılımı tespit edilememektedir ve bu sonsal dağılım iterasyon ile elde edilebilir. Aşağıda, Gibbs örnekleme ile Tanner ve Wong'un DA yöntemlerinin özel bir durumu için yaptıkları algoritma (3, 35, 45) sunulmaktadır.

Sonsal dağılım, eksik verinin rassal değerini, parametreler için karşılıklı olarak değiştirerek türetilmeye çalışılmaktadır.  $t$  iterasyonda aşağıdaki adımlar incelenir.

- Atama adımı; parametrenin o andaki değeri  $\theta^{(t)}$  verildiğinde  $Y_{\text{Eksik}}^{(t+1)}$  bunun koşullu tahmin dağılımı  $P[Y_{\text{Eksik}} | Y_{\text{Gözlem}}, \theta^{(t)}]$ 'dan elde edilir.
- Sonsal adım;  $Y_{\text{Eksik}}^{(t+1)}$  verildiğinde  $\theta^{(t+1)}$ 'in tam veri sonsal  $P[\theta | Y_{\text{Gözlem}}, Y_{\text{Eksik}}^{(t+1)}]$ 'dan elde edilir.

Birbirinden bağımsız ve aynı dağılıma sahip örneklemler için, eksik çokdeğişkenli veri seti için aşağıdaki eşitliğin doğru olduğu söylenilebilir.

$$P[Y_{\text{Eksik}} | Y_{\text{Gözlem}}, \theta] = \prod_{i=1}^n P[y_{i(\text{Eksik})} | y_{i(\text{Gözlem})}, \theta] \quad (2.30)$$



Burada  $y_{i(\text{Eksik})}$ , i. satırdaki eksik veriyi ve  $y_{i(\text{Gözlem})}$  ise i. satırdaki gözlenen veriyi temsil etmektedir. Dolayısıyla yukarıda tanımlanan ilk adımda eksik veri satırlar birbirinden bağımsız olacak şekilde atanmaktadır (40)

Yukarıda sözü edilen atama yöntemlerinden başka tam bilgi en büyük benzerlik tahmincisi (full information maximum likelihood estimation), yapısal denklem modelleme yaklaşımı (structural equation modeling approach) ve örüntü karışımı model yaklaşımı (pattern mixture model approach) diğer atama yöntemlerinden bazılarıdır (27).

## 2.6- Little'in TROK (MCAR) Testi

Roderick J. A. Little'in ki-kare istatistiği, verilerin tamamıyla rassal olarak eksik (TROK) olup olmadığını test eder. Bu test yardımıyla, tüm veri setindeki eksik veri yapısı analiz edilir ve rasgele eksik veri süreci için, beklenen yapı ile karşılaştırılır. Eğer anlamlı farklılıklar yok ise eksik veri TROK olarak adlandırılır (2).

Burada, test istatistiği hesaplandıktan sonra elde edilen P olasılık değeri 0.05 değerinden küçük ( $P < 0.05$ ) olduğunda, veri yapısının TROK olmadığı sonucuna varılır (13).

İstenen ki-kare test istatistiği,

$$\chi^2_{\text{MCAR}} = \sum_{\text{herbir tek yapı}} (\text{yapıdaki birim sayısı}) * (\bar{X}^E \text{ den elde edilen yapı ortalamasının Mahalanobis } D^2 \text{ uzaklığı}) \quad (2.31)$$

formülüyle hesaplanır. İlgili ki-kare test istatistiği serbestlik derecesi,

$$SD_{\text{MCAR}} = \sum_{\text{herbir tek yapı}} (\text{eksik olmayan değişken sayısı}) - v \quad (2.32)$$

olarak alınır. Analiz sonucunda verinin TROK olmadığı kararı verilirse, eksik veri için kullanılan bazı yöntemler, örneğin varyans, yanlı ve tutarsız tahminler verecektir (13).

TROK mekanizması için ayrıca Little'in  $d^2$  değeri de kullanılabilir. Bu değer normalleştirilmiş Mahalanobis uzaklığı değeridir. Mahalanobis uzaklığı hesaplanırken genel ortalama ile değişkenlerde aynı eksik veri yapısına sahip, alt veri gruplarının

ortalamları arasındaki uzaklıklar hesaplanır. Burada kullanılan uzaklığa normalleştirilmiş denilmesinin nedeni, kullanılan kovaryans matrisinin her bir eksik veri yapısı (patterni) örneklem ortalaması için kovaryans matrisi olmasıdır, tek bir gözlem matrisi, pattern de bulunan gözlem birimi sayısına bölünmektedir (12, 22, 23, 35).

Eksik veri sürecinin, rasgele ortaya çıkıp çıkmadığının araştırılmasında kullanılan diğer bir yöntem ise t testi ile yapılır. Burada veri yapısındaki bir değişkene ilişkin gözlemler eksik veri içerenler ve içermeyenler olarak iki gruba ayrılarak ilgilenilen diğer değişkenleri açısından bu iki grup arasında anlamlı fark olup olmadığı araştırılır. Bu araştırma, bağımsız gruplarda iki ortalama arasındaki farkın anlamlılık testi (t testi) ile yapılır. Anlamlı fark, rasgele olmayan eksik veri sürecinin var olabileceğini gösterir. Ancak, bazı anlamlı farkların rasgele ortaya çıkması sözkonusu olabilir. Dolayısıyla, rasgele olmayan bir yapıdan söz edebilmek için bir dizi farklılığın ortaya çıkması gerektiği unutulmamalıdır (2).

Rasgeleliğin araştırılmasında kullanılan bir başka yöntem ise, değişkenlerin eksik ve eksik olmayan şeklinde iki kategoriye indirilmesi ve değişken çiftleri arasında Pearson korelasyon katsayısının hesaplanmasıdır. Bu amaçla eksik olmayan tam veriler 1, eksik veriler 0 olarak yeniden kodlanır. Bulunan korelasyon katsayıları, her bir değişken çifti için eksik veriler arasındaki ilişki miktarının derecesini bildirir (Veriler 0 ve 1 şeklinde dönüştürüldüğünde, Pearson korelasyon katsayısı (r), ilgili değişkenlerle elde edilecek çapraz tablo yardımıyla bulunacak Phi katsayısına eşit çıkar). Buna göre, düşük korelasyonlar her bir değişken çifti için eksik verideki rasgeleliği gösterir. Ancak, rasgele olmayan eksik veri sürecini tanımlamak için kesin bir kural söz konusu değildir. Diğer taraftan korelasyon katsayısının anlamlılık testleri ise rasgeleliğin derecesini tanımlamada oldukça tutucu bir kestirim sağlamaktadır. Eğer rasgelelik tüm değişken çiftleri için sözkonusu ise, araştırmacı, veri mekanizmasını TROK (MCAR) olarak belirler. Ancak, bazı değişken çiftleri arasında anlamlı korelasyonlar var ise veri rasgele eksik (ROK, MAR) olarak düşünülebilir (2, 22).

### 3- GEREÇ VE YÖNTEMLER

#### 3.1- Veri Türetimi

Araştırmada türetilmiş verilerden yararlanılmıştır. Veri türetimi belirli koşullar göz önüne alınarak bu koşullara uygun biçimde S-Plus paket programı kullanılarak yapılmıştır.

Veri setleri; birim sayısı sırasıyla 50, 100, 200, 300, 400 ve 500 birim içerecek şekilde, çokdeğişkenli standart normal dağılım (SND) varsayımından yararlanılarak türetilmiştir. Her bir veri setinde  $N \sim (0,1)$  parametrelili 4 değişken ( $v=4$ ) türetilmiş ve bunlar sırasıyla X1, X2, X3, ve X4 olarak isimlendirilmiştir. Değişken sayısının dört ile sınırlandırılması analizlerin çabuklaştırılması ve çok değişik kombinasyonlarda yöntemlerin etkinliklerini araştırmaya yöneliktir. Veri setlerinde X1, X2, X3 ve X4 değişkenlerinin aralarındaki korelasyonlar düşük ( $-0.30 < r < 0.30$ ) ve yüksek ( $r < -0.70$  veya  $r > 0.70$ ) olacak şekilde türetilmiştir. Bu korelasyonların önemliliği üzerinde durulmamıştır. Ayrıca eksik veri analizinde grup etkisini test edebilmek için ikili kategoriye sahip (0,1) ve tekdüze (uniform) dağılımdan türetilmiş X5 değişkeni veri matrisinin sonuna  $v=5$  olacak şekilde eklenmiştir (45).

$N=50,100,200,300,400,500$  olacak şekilde rasgele türetilen tam veri setlerinde bulunan her bir değişkendeki gözlemler sayıları, S-Plus paket programı kullanılarak, rasgele koşullar altında sırasıyla %5, %10, %15 ve %20 azaltılmış ve eksik veri analizine esas olacak eksik veri setleri elde edilmiştir (45).

Tam veri setlerinin eksiltilmesi aşamasında iadeli seçim uygulanmıştır. S-Plus paket programında, rasgele koşullar altında uygulanan eksiltme aşamasında bir değişkene ait herhangi bir gözlem değeri veri setinden çıkartıldıktan sonra bu değer tekrar örnekleme çıkabilmektedir. Örneğin; 15. gözlem biriminde, X1 değişkenine ait gözlem değeri ilk aşamada eksiltilmiş ise, ilerleyen eksiltme aşamalarında aynı birime ait gözlem değeri tekrar örnekleme çıkabilmekte ve bu durumda da ilgilenilen gözlem değeri eksik olduğundan dolayı buradaki değer için eksiltme işlemi yapılamamaktadır. Böylece, 100 birim içeren bir veri setinde, eksiltme aşamasında böyle bir durumla

karşılaştırılmış ise, aslında %10 eksik olması gereken veri seti, daha az azaltılmış (örneğin %9 gibi) olabilmektedir. Böylece, değişik birim sayıları içeren düşük ve yüksek korelasyonlara sahip veri setlerinin eksiltilmesinde adı geçen %5, %10, %15 veya %20 eksiltme işlemleri yaklaşık değerler olarak gerçekleşmiştir. Yerine koyma yöntemi kullanılması veri setlerindeki değer yerleşimlerinin tamamen rasgelelik koşullarını taşımasını sağlamak içindir (45).

Düşük ve yüksek korelasyonlara sahip 50, 100, 200, 300, 400 ve 500 birimden oluşan toplam 12 veri seti içerisindeki değişkenler sırasıyla %5, %10, %15 ve %20 gözlem eksiltilmiş ve eksik birimlere sahip 48 veri seti elde edilmiştir. Bu değişik birim sayılarına ve farklı korelasyonlu değişkenlere sahip 48 veri seti, kayıp veri analizlerin uygulanabilmesi amacıyla öncelikle TROK (MCAR) testine tabi tutulmuş, eğer veri seti TROK mekanizmasına sahip ise atama yöntemleri yardımıyla eksik veriler uygun değerlerle doldurulmuştur. Eksik veri tamamlama yöntemleri ile ayrı ayrı tamamlanan veri setleri, orijinal türetilmiş (**tam**) ve değişik oranlarda eksiltilmiş (**eksik**) veri setlerinin parametreleri hesaplanmış ve bu parametreler birbirleriyle benzerlikleri bakımından karşılaştırılmıştır. Eksik veri analizi yöntemlerine göre doldurulmuş veri setlerinin orijinal veri seti parametrelerine yaklaşımları analiz edilmiş ve yöntemlerin etkinliği benzerliklere/yakınlıklara göre belirlenmiştir.

### **3.2- Eksik Veri Tamamlama Yöntemleri**

Tam ve değişik oranlarda eksiltilmiş farklı iki korelasyon düzeyine göre türetilmiş ek veri setlerinin analizinde **birimin silinmesi** (listwise deletion), eksik değerli değişkenlerden değerlendirmeye alınacak karşılıklı verileri tamam birimlerin alınması (**karşılıklı silme**, pairwise deletion), veri yapısına uygun olarak eksik verinin yerine tahmini değerlerin doldurulması ile ilgili **Ortalama Atama**, **EM** ve **Regresyon Atama** yöntemlerinden yararlanılmıştır. Genel bilgiler bölümünde sözü edilen diğer bazı yöntemler türetilen veri setlerimizin taşıdığı koşullar bakımından uygun olmadıkları için araştırmada kullanılmamıştır.

### 3.2.1- Tam Gözlemlerin Kullanılması Yöntemi (Listwise Method)

Bu yöntemde, yalnızca eksik veri içermeyen (complete) birimlerin aldıkları değerler kullanılmış ve parametre tahminleri aşağıdaki formüllere göre hesaplanmıştır.

Formüllerde j ve k indisleri kantitatif değişkenleri göstermektedir.

Liste bazında silme yöntemi ile ortalama;

$$\bar{x}^L = [\bar{x}_j^L] = \left[ \sum_i x_{ij} / n_c; \quad i \in I(J) \right] \quad (3.1)$$

formülüyle hesaplanır. Kovaryans ( $C^L$ ) ve korelasyon ( $R^L$ ) değerleri ise;

$$C^L = [c_{jk}^L] = \left[ \sum_i (x_{ij} - \bar{x}_j^L) * (x_{ik} - \bar{x}_k^L) / (n_c - 1); \quad i \in I(J) \right] \quad (3.2)$$

$$R^L = [r_{jk}^L] = [c_{jk}^L / (c_{jj}^L * c_{kk}^L)^{1/2}] \quad (3.3)$$

formülleri yardımıyla hesaplanmıştır (13, 22).

### 3.2.2- Eldeki Tüm Bilginin Kullanılması (Karşılıklı Silme, Pairwise Method)

Bu yöntemde değerleri karşılıklı olarak tamam olan birimlerin analize alınması biçiminde uygulanmaktadır. Bu yöntemde aşama aşama her bir değişkenin ortalama ve varyansları ayrı ayrı hesaplanır, ayrıca çift olarak ele alındıklarında ise ulaşılabilen tüm çiftlerin kovaryansları hesaplanır (7).

Karşılıklı Silme yönteminde parametre tahminleri aşağıdaki formüllere göre hesaplanmıştır. Bu formüllerde j ve k indisleri kantitatif değişkenleri ve l indisi ise tüm değişkenleri göstermektedir.

Ortalama;

$$\bar{x}^P = [\bar{x}_{lk}^P] = \left[ \sum_i x_{ik} / n_{lk}; \quad i \in I(1, k) \right] \quad \text{formülü ile} \quad (3.4)$$

Standart sapma

$$\hat{\sigma}^P = [\hat{\sigma}_{lk}^P] = \left[ \left( \sum_i (x_{ik} - \bar{x}_{lk}^P)^2 / (n_{lk} - 1) \right)^{1/2}; \quad i \in I(l, k) \right] \text{ formülü ile} \quad (3.5)$$

hesaplanır.

Kovaryans ( $C^P$ ) ve korelasyonlar ( $R^P$ ) aşağıdaki gibi hesaplanır..

$$C^P = [c_{jk}^P] = \left[ \sum_i (x_{ik} - \bar{x}_{jk}^P) * (x_{ij} - \bar{x}_{kj}^P) / (n_{jk} - 1); \quad i \in I(j, k) \right] \quad (3.6)$$

$$R^P = [r_{jk}^P] = [c_{jk}^P / (\hat{\sigma}_{jk}^P * \hat{\sigma}_{kj}^P)] \quad (3.7)$$

Bu yöntemin ayırt edici özelliği; her bir korelasyon için ayrı gözlem çiftlerinin ve her hesaplamada farklı sayıda gözlemin kullanılmasıdır. Bu yöntem yardımıyla elde edilen korelasyonlar bütün örneklemin temsilcisi olarak kabul edilir ve değer ataması yapılmaz.

### **3.2.3- Değer Atama Yöntemleri (Imputation Methods)**

TROK (MCAR) mekanizmasının geçerli olduğu eksik veri setlerinde eksik verinin giderilmesine yönelik olarak başvuru yapılan yöntemler ortalama atama, regresyon atama ve EM atama yöntemleridir.

Değer atama yöntemleri; örneklemin tümüne genellenmek amacıyla gözlemlerin bir alt kümesinden elde edilen tüm bilginin kullanılması ya da eksik veri yerine atanacak değerlerin kestirilmesi şeklinde tanımlanabilirler.

#### **3.2.3.1- Ortalama Atama Yöntemi (Mean Substitution)**

Bu yöntem eksik değer içeren değişken için eksik değerli değişken verilerinin ortalamasının yerleştirilmesi biçiminde uygulanmaktadır. Bu işlem sonucunda, tam ve atanmış veriler için elde edilen ortalama, tam verilerle elde edilen ortalamaya eşit çıkmaktadır.

### 3.2.3.2- Regresyon Atama (Regression Imputation) Yöntemi

Eksik veri içeren veri setlerinde değişken sayısına göre basit ya da çoklu regresyon yöntemi kullanılarak eksik veriler tahmin edilmiş ve eksik veriler yerine atanmıştır.

Hangi regresyon modelinin kullanılacağı eksik verilerin dağılımı incelenerek belirlenmiştir.

Regresyon atama yönteminde eksik verilerin tahmininde aşağıdaki formül ve gösterimlerden yararlanılmaktadır. Bu formüllerde j ve k indisleri kantitatif değişkenler için, l indisi ise tahmini değişkenler için kullanılmıştır.

$$x_{ij}^R = \begin{cases} x_{ij} & \text{eğer } x_{ij} \text{ kayıp değilse} \\ \text{regresyon tahmini } x_{ij} & \text{eğer } x_{ij} \text{ kayıpsa} \end{cases}$$

Regresyon tahmini  $x_{ij}$

$$x_{ij}^R = \beta_{0,ij} + \sum_1 \beta_{1,ij} * x_{il} + \varepsilon_{ij} \quad l \in J_1 = J(l : x_{il} \text{ kayıp değilse ve } l \neq j) \quad (3.8)$$

formülüyle hesaplanır. Burada;

$[\beta_{0,ij}, \beta_{1,ij}]$  ;  $\text{Diag}(\bar{X}^P) = [\bar{x}_{jj}^P]$ 'dan,  $C^P$ 'nin  $J_1$  köşegenlerinin en iyi (best) q dönüşümü yapılarak hesaplanır. Buradaki en iyi, ileriye doğru aşamalı seçimdir (forward stepwise selection). q ise maksimum bağımsız değişken sayısına eşit ya da küçük bir değere sahiptir. Aynı zamanda q değeri, belli bir F değeri belirlenerek de sınırlandırılabilir.

$\varepsilon_{ij}$  , rassal hata terimidir.

Her bir eksik  $x_{ij}$  değeri için regresyon katsayıları kümesi  $(\beta_{0,ij}, \beta_{1,ij})$  ve  $\varepsilon_{ij}$  hesaplanır.

Burada ortalama;

$$\bar{x}^R = [\bar{x}_j^R] = \left[ \sum_i x_{ij}^R / n; \quad i \in I \right] \quad (3.9)$$

formülüyle hesaplanır. Değişkenlere ilişkin kovaryans;

$$C^R = [c_{jk}^R] = \left[ \sum_i (x_{ij}^R - \bar{x}_j^R) * (x_{ij}^R - \bar{x}_j^R) / (n-1); \quad i \in I \right] \quad (3.10)$$

formülü, korelasyonlar ise;

$$R^R = [r_{jk}^R] = [c_{jk}^R / (c_{jj}^R * c_{kk}^R)^{1/2}] \quad (3.11)$$

formülü yardımıyla hesaplanır.

### 3.2.3.3- EM (Expectation-Maximization) Yöntemi

EM algoritması eksik verilere sahip birimlerin tahmininde kullanılan maksimum benzerlik tahminlerini içeren bir yöntemdir. EM yöntemi, yinelemeli (iterative) ve iki aşamalı bir yöntem olarak uygulanmış, E aşamasında eksik veriler için en iyi olası kestirimler bulunmuş, M aşamasında ise eksik veri atandığında veri setinin ortalama, standart sapma ya da korelasyona ilişkin kestirimleri elde edilmiştir. Bu süreç kestirilen değerlerdeki değişimin önemsenmeyecek derecede azalmasına kadar devam ettirilmiştir.

EM atama yönteminde kullanılan formül ve gösterimler aşağıdaki gibidir. Bu formüllerde j ve k indisleri kantitatif değişkenleri ve l indisi ise tahmini değişkenleri göstermektedir.

Eksik verilerin tahminleri, ortalama vektörü ve kovaryans matrisi;

$$x_0 = [\bar{x}_j^0] = \text{Diag}(\bar{x}^P) = [\bar{x}_{jj}^P] \quad (3.12)$$

$$C_0 = [c_{jk}^0] = C^P = [c_{jk}^P] \quad (3.13)$$

formülleriyle hesaplanır.



Burada m=1 den M'ye kadar veya yakınsama elde edilinceye kadar. Eğer  $x_{ij}$  gözlem değeri eksik değilse (tam ise)

$$x_{ij}^m = x_{ij} \quad (3.14)$$

eğer  $x_{ij}$  eksikse m. İterasyon sonucu tahmin edilen değer

$$x_{ij}^m = \beta_{0,ij}^{m-1} + \sum_l \beta_{l,ij}^{m-1} * x_{il}; \quad l \in J_2 = J(1: x_{il} \text{ kayıp değilse ve } l \neq j) \quad (3.15)$$

formülü yardımıyla hesaplanır. Burada  $[\beta_{0,ij}^{m-1}, \beta_{l,ij}^{m-1}]$  değerleri  $\bar{x}_{m-1}$  ve  $C_{m-1}$ 'den hesaplanır.  $\bar{x}_m$  ve  $C_m$  aşağıdaki gibi hesaplanır.

$$\bar{x}_m = [\bar{x}_j^m] = \left[ \sum_i w_i * x_{ij}^m / \sum_i w_i; \quad i \in I \right] \quad (3.16)$$

$$C_m = [c_{jk}^m] = \left[ \frac{\sum_i w_i * x_{ij}^m (x_{ij}^m - x_j^{-m}) * (x_{ik}^m - x_k^{-m}) + \sum_i \sum_s c_{j,s|J_2}^{m-1}}{(n-1) * \sum_i w_i / n}; \quad i \in J_2, s \notin J_2 \text{ ve } s \neq j \right] \quad (3.17)$$

burada  $c_{j,s|J_2}^{m-1}$ ,  $C_{m-1}$  değerini pivot eden (döndüren)  $J_2$  nin j. satır s. elemanıdır.

$$w_i = \begin{cases} 1 & \text{çokdeğişkenli normal için} \\ \frac{1 - \alpha + \alpha * \lambda^{1+p/2} * \exp((1-\lambda) * D^2 / 2)}{1 - \alpha + \alpha * \lambda^{p/2} * \exp((1-\lambda) * D^2 / 2)} & \text{bozulmuş normal için} \\ (sd + p) / (sd + D^2) & \text{t(sd) dağılımı için} \end{cases} \quad (3.18)$$

$\alpha$  = Bozulma oranı

$\lambda$  = Standart sapma oranı

p = Bağımsız değişkenlerin (tahmincilerin) sayısı =  $J_2$  deki indislerin sayısı

$D^2$  = Birimin ortalamadan olan Mahalanobis uzaklığının karesi

$$= \sum_{jk} (x_{ij}^m - \bar{x}_j^m) * (c_{jk}^m)^{-1} * (x_{ik}^m - \bar{x}_k^m) \quad (3.19)$$

Yukarıdaki formülde  $(c_{jk}^m)^{-1}$  ifadesi  $C_m^{-1}$ 'in  $jk$ . elemanıdır.

Algoritmanın sona ermesi için yakınsama değeri,

$$|c_{jj}^m - c_{jj}^{m-1}| / c_{jj}^m \leq \text{Yakınsama} \quad \text{olarak alınmıştır.} \quad (3.20)$$

Veri atama (doldurma);

$$X_i^E = [x_{ij}^E] = [x_{ij}^{m'}] \quad (3.21)$$

burada  $m'$ ,  $m$ 'nin son değeridir.

Ortalama ve kovaryanslara ilişkin formüller aşağıdaki gibidir.

$$\bar{x}^E = [\bar{x}_j^E] = \bar{x}_{m'} = [\bar{x}_j^{m'}] \quad (3.22)$$

$$C^E = [c_{jk}^E] = C_{m'} = [c_{jk}^{m'}] \quad (3.23)$$

Korelasyon ise,

$$R^E = [r_{jk}^E] = [c_{jk}^E / (c_{jj}^E * c_{kk}^E)^{1/2}] \quad (3.24)$$

formülü yardımıyla hesaplanır (3, 13, 22).

### **3.3- Little'in TROK (MCAR) Testi**

Eksik verilerin mekanizmasının TROK (MCAR)'a uygunluğunu test etmek için Little MCAR testi kullanılmıştır. Hesaplanan Little'in ki-kare istatistiği, önemlilik göstermiyorsa verilerin TROK mekanizmasında olduğu kabul edilir. Diğer halde ROK ya da IEM mekanizmasının verileri temsil edebileceğine karar verilir.

Little ki-kare test istatistiği,

$$\chi_{MCAR}^2 = \sum_{\text{herbir tek yapı}} (\text{yapıdaki birim sayısı}) * (\bar{x}^E \text{ den elde edilen yapı ortalamasının Mahalanobis } D^2 \text{ uzaklığı}) \quad (3.25)$$

formülüyle hesaplanmış, Little ki-kare test istatistiğinin önemliliği;

$$SD_{MCAR} = \sum_{\text{herbir tek yapı}} (\text{eksik olmayan deęişken sayısı}) - v \quad (3.26)$$

serbestlik dereceli kikare daęılımının kritik deęerlerine göre belirlenmiştir.

Little kikare test istatistięi, Little  $d^2$  ve t testine göre daha güçlü olduęu için kullanılmıştır (13).

Deęişik veri setlerindeki ortalamaların benzerliklerini test etmek için t testinden yararlanılmıştır.

Ortak varyans yaklaşımının geçerli olduęu durumlar için Ortak varyans t testi (equal variance assumed t test), farklı varyans durumunda ise farklı varyans t testi (seperate variance assumed t test) uygulanmıştır.

t test istatistięi ařaęıdaki formül yardımıyla hesaplanmıştır.

$$t_{jk} = \frac{\bar{X}_{jk}^p - \bar{X}_{k|j \text{ deęişkeni kayıpsa}}}{\left( \frac{\hat{\sigma}_{jk}^p}{n_{jk}} + \frac{\hat{\sigma}_{k|j \text{ deęişkeni kayıpsa}}}{n_{kk} - n_{jk}} \right)^{1/2}} \quad (3.27)$$

Farklı varyans yaklaşımında serbestlik derecesi ařaęıdaki gibi belirlenir.

$$sd_{jk} = \frac{\left( \frac{\hat{\sigma}_{jk}^p}{n_{jk}} + \frac{\hat{\sigma}_{k|j \text{ deęişkeni kayıpsa}}}{n_{kk} - n_{jk}} \right)^2}{\frac{(\hat{\sigma}_{jk}^p)^2}{n_{jk} - 1} + \frac{(\hat{\sigma}_{k|j \text{ deęişkeni kayıpsa}})^2}{n_{kk} - n_{jk} - 1}} p(2 \text{ yönlü})_{jk} = 1 - 2 * |0.5 - tcdf(t_{jk}, sd_{jk})| \quad (3.28)$$

Formülde 'tcdf' (t cumulative distribution function), yığılımlı t daęılım fonksiyonunu temsil etmektedir.

### 3.4- Eksik Veri Analizinde Belirtici İstatistikler

Tekdeğişkenli ve çokdeğişkenli veri setlerinde tanımlayıcı ve diğer istatistikler aşağıdaki verilen formüllerden veri seti için uygun olanları kullanılarak hesaplanmıştır.

Aşağıdaki formülde,  $\bar{x}$ , değişkenlere ilişkin ortalama vektörünü göstermektedir. Burada değişkenler için j indeksi kullanılmıştır.

$$\bar{x} = [\bar{x}_j] = \left[ \sum_i x_{ij} / n_j; i \in I(j) \right] \quad (3.29)$$

Değişkenlere ait standart sapma değerleri,

$$\hat{\sigma} = [\hat{\sigma}_j] = \left[ \sum_i (x_{ij} - \bar{x}_j)^2 / (n_j - 1)^{1/2}; i \in I(j) \right] \quad (3.30)$$

şeklinde hesaplanmaktadır.

Veri setindeki aşırı küçük değerler (Extremely low)  $NL = [nl_j] = [x_{ij} \text{ değerlerinin sayısı} < \text{alt\_limit}_j]$  şeklinde, aşırı yüksek değerler ise (Extremely high),  $NH = [nh_j] = [x_{ij} \text{ değerlerinin sayısı} > \text{üst\_limit}_j]$  şeklinde hesaplanmaktadır. Formüllerde ;

$$\text{alt\_limit}_j = \begin{cases} \bar{x}_j - 2 * \hat{\sigma}_j & \text{eğer } v * n * \log_{10}(n) > 150.000 \\ \text{j. değişkenin \%25'i} & \text{eğer } v * n * \log_{10}(n) \leq 150.000 \end{cases}$$

ve

$$\text{üst\_limit}_j = \begin{cases} \bar{x}_j + 2 * \hat{\sigma}_j & \text{eğer } v * n * \log_{10}(n) > 150.000 \\ \text{j. değişkenin \%75'i} & \text{eğer } v * n * \log_{10}(n) \leq 150.000 \end{cases}$$

olarak kullanılmıştır (13).

#### 4- BULGULAR

Türetilen n=50, 100, 200, 300, 400, 500 birimlik yüksek ve düşük korelasyonlu 5 değişkenli veri setlerinde eksik veri yüzdelere göre TROK özelliği taşıyıp taşımadıklarına ilişkin Little  $\chi^2$  testi sonuçları Çizelge 4.1’de verilmiştir.

Çizelge 4.1: Türetilen veri setlerine ait Little’nin TROK (MCAR) test sonuçları

Veri setleri	Korelasyon Düzeyleri	Eksik olma durumu (%)	$\chi^2$	sd	P
50 birimlik Veri setleri	Düşük Korelasyonlu	5	15.36	12	0.22
		10	14.13	18	0.72
		15	3.44	17	0.99
		20	11.09	22	0.97
	Yüksek Korelasyonlu	5	5.32	14	0.98
		10	7.74	18	0.98
		15	4.02	19	0.99
		20	11.77	20	0.92
100 birimlik Veri setleri	Düşük Korelasyonlu	5	7.13	17	0.98
		10	10.24	19	0.95
		15	4.55	21	0.99
		20	12.64	24	0.97
	Yüksek Korelasyonlu	5	13.46	14	0.49
		10	12.91	20	0.88
		15	12.54	23	0.96
		20	39.34	27	0.06
200 birimlik Veri setleri	Düşük Korelasyonlu	5	16.06	17	0.52
		10	24.43	20	0.22
		15	18.47	25	0.82
		20	18.27	26	0.87
	Yüksek Korelasyonlu	5	1.43	14	0.99
		10	28.89	22	0.15
		15	4.66	25	0.99
		20	7.78	26	0.99
300 birimlik Veri setleri	Düşük Korelasyonlu	5	4.09	16	0.99
		10	12.32	25	0.98
		15	7.99	26	0.99
		20	12.68	27	0.99
	Yüksek Korelasyonlu	5	7.25	18	0.99
		10	3.33	24	0.99
		15	13.94	25	0.96
		20	15.88	27	0.96
400 birimlik Veri setleri	Düşük Korelasyonlu	5	21.96	20	0.34
		10	15.93	24	0.89
		15	14.53	25	0.95
		20	28.23	28	0.45
	Yüksek Korelasyonlu	5	29.85	22	0.12
		10	27.01	25	0.36
		15	29.83	28	0.37
		20	38.07	26	0.06
500 birimlik Veri setleri	Düşük Korelasyonlu	5	20.92	21	0.46
		10	28.70	24	0.23
		15	14.84	26	0.96
		20	28.28	28	0.45
	Yüksek Korelasyonlu	5	17.25	20	0.64
		10	20.06	25	0.74
		15	17.49	26	0.89
		20	9.89	28	0.99

Çizelge 4.1'den de görüleceği gibi tüm veri setlerindeki eksik olma durumları TROK yapıdadır. Hesaplanan tüm Little  $\chi^2$  istatistiklerine ait olasılık değerleri kritik değer olan 0.05 ( $P>0.05^{ns}$ ) değerinden büyük değerlere sahiptir. Tüm veri setleri TROK mekanizmasına sahiptir.

#### **4.1- 50 Birimlik Türetilmiş Veri Setleri**

Düşük ve yüksek korelasyonlara sahip 50 birimlik veri setlerine ilişkin analizler aşağıdaki bölümlerde verilmiştir.

##### **4.1.1- Düşük Korelasyonlu 50 Birimlik Veri Seti**

50 birim içeren tam veri setine ilişkin değişkenler arasındaki korelasyonlar Çizelge 4.2'de verilmiştir. Çizelge 4.2 incelendiğinde değişkenler arasındaki korelasyon katsayıları 0.40'dan daha düşüktür ( $r<0.40$ ). Ancak değişkenlerden bazıları arasındaki bu korelasyonlar önemli olarak saptanmıştır.

Çizelge 4.2- 50 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri

	X1	X2	X3	X4
X1	1	0.05 ( $P>0.05$ )	0.32 ( $P<0.05$ )	0.29 ( $P>0.05$ )
X2		1	0.24 ( $P>0.05$ )	0.32 ( $P<0.05$ )
X3			1	0.40 ( $P<0.001$ )
X4				1

Not: Çizelgede parantez içindeki değerler olasılık değerleridir.

Çizelge 4.3'te 50 birimden oluşan düşük korelasyonlara sahip tam veri setinde sırasıyla %5, %10, %15 ve %20 değer eksiltilerek elde edilen veri setlerindeki eksik olma sayı ve yüzdeleri verilmiştir. Çizelge 4.3 incelendiğinde eksiltmelerde verilen yüzde değerlerine yaklaşık azaltmalar gerçekleşmiş ve hedeflenen eksiltmelere ulaşıldığı kabul edilmiştir.

Çizelge 4.3- 50 birim içeren düşük korelasyonlu değişkenlere sahip veri setinde. değişkenlerin eksik olma sayı ve yüzdeleri

Değişken	Tam veri (n)	%5 Eksik		%10 Eksik		%15 Eksik		%20 Eksik	
		sayı	%	sayı	%	sayı	%	sayı	%
X1	50	3	6	4	8	8	16	9	18
X2	50	3	6	5	10	8	16	9	18
X3	50	3	6	5	10	8	16	8	16
X4	50	3	6	5	10	8	16	9	18
X5	50	0	0	0	0	0	0	0	0

Çizelge 4.4'te tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen ortalama ve standart sapma değerleri verilmiştir.

Çizelge 4.4- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri

		X1 ( $\bar{X} \pm S$ )	X2 ( $\bar{X} \pm S$ )	X3 ( $\bar{X} \pm S$ )	X4 ( $\bar{X} \pm S$ )
<b>Tam Veri</b>		<b>0.377±1.184</b>	<b>0.247±1.002</b>	<b>-0.034±1.200</b>	<b>0.009±0.928</b>
%5 eksik	Listwise	0.386±1.128	0.262±1.054	-0.057±1.165	-0.001±0.953
	Pairwise	0.316±1.190	0.232±0.995	-0.093±1.158	-0.017±0.932
	Ortalama	0.316±1.153	0.232±0.965	-0.093±1.122	-0.017±0.903
	EM	0.319±1.188	0.224±0.991	-0.093±1.155	-0.029±0.934
	Regresyon	0.387±1.128	0.262±1.054	-0.057±1.165	-0.001±0.953
%10 eksik	Listwise	0.401±1.092	0.264±0.999	-0.181±1.169	0.138±0.973
	Pairwise	0.410±1.081	0.247±1.025	-0.040±1.162	-0.011±0.966
	Ortalama	0.410±1.036	0.247±0.971	-0.040±1.101	-0.011±0.915
	EM	0.408±1.077	0.247±1.021	-0.071±1.183	0.025±0.997
	Regresyon	0.401±1.092	0.264±0.998	-0.181±1.169	0.138±0.973
%15 eksik	Listwise	0.274±1.265	0.408±1.084	0.112±1.003	0.058±0.925
	Pairwise	0.288±1.215	0.359±1.001	-0.014±1.201	-0.049±0.922
	Ortalama	0.288±1.112	0.359±0.916	-0.014±1.098	-0.049±0.843
	EM	0.284±1.217	0.332±1.001	-0.003±1.200	-0.030±0.924
	Regresyon	0.284±1.221	0.388±0.994	-0.129±1.151	-0.092±0.900
%20 eksik	Listwise	0.032±1.386	0.432±1.056	-0.320±1.120	-0.215±0.945
	Pairwise	0.360±1.268	0.172±1.005	-0.080±1.181	-0.097±0.820
	Ortalama	0.360±1.145	0.172±0.908	-0.080±1.081	-0.097±0.741
	EM	0.345±1.268	0.205±1.004	-0.045±1.174	-0.091±0.821
	Regresyon	0.244±1.289	0.253±1.012	-0.180±1.177	-0.076±0.787

Düşük korelasyonlu değişkenlere sahip 50 birimlik veri setlerinin analizleri sonucunda; %5'lik eksik veri setinde, Listwise ve Regresyon atama yöntemleri tam veri setinden elde edilen belirtici istatistikler t testi ile karşılaştırılmış ve benzer sonuçlar vermiştir ( $P > 0.05$ ). %10'luk eksik veri setinde, aynı dağılım yapısına sahip 4 farklı değişken olmasına rağmen tüm atama yöntemlerinden farklı sonuçlar elde edilmiştir ( $P < 0.05$ ). Genel olarak X2, X3 ve X4 değişkenlerinde benzer ortalamalar Pairwise ve Ortalama atama sonucunda oluşan veri setinden elde edilmiştir. %15'lik eksik veri

setinde, X2, X3 ve X4 deęişkenlerinde EM atama yöntemi, tam veri setinin ortalamalarına daha yakın bulunurken, X1 deęişkeninde Pairwise ve Ortalama atama yöntemi, tam veri setine daha yakın sonuçlar vermiş ancak bu ortalamalar yine de birbirinden oldukça farklılık göstermiştir ( $P<0.05$ ). %20'lik eksik veri setinde, farklı atama yöntemleri ile tam veri setinden elde edilen belirtici istatistikler arasında benzerlik gözlenmemiştir. Türetilen 4 deęişkende de bu sonuçlar benzerlik göstermemektedir ( $P<0.05$ ).

Çizelge 4.5'te tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 deęişkeni bağımlı ve X2, X3 ve X4 deęişkenleri bağımsız deęişkenler olarak alındığında, elde edilen regresyon katsayıları, bu katsayıların anlamlılık düzeyleri ve belirleyicilik katsayıları ( $R^2$ ) verilmiştir.

Çizelge 4.5- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları

		Sabit ( $b_0$ )	$b_1$	$b_2$	$b_3$	$R^2$
	<b>Tam Veri</b>	<b>0.406*</b>	<b>-0.096</b>	<b>0.253</b>	<b>0.275</b>	<b>0.142</b>
%5 eksik	Listwise	0.392*	0.033	0.257	0.147	0.128
	Ortalama	0.350*	-0.040	0.249	0.182	0.099
	EM	0.361*	-0.060	0.244	0.207	0.110
	Regresyon	0.277	-0.063	0.229	0.081	0.083
%10 eksik	Listwise	0.418*	-0.050	0.193	0.217	0.109
	Ortalama	0.424**	-0.010	0.213	0.247	0.132
	EM	0.438**	-0.070	0.267	0.235	0.172
	Regresyon	0.439**	-0.050	0.196	0.175	0.090
%15 eksik	Listwise	0.243	-0.024	0.189	0.333	0.102
	Ortalama	0.316	-0.040	0.073	0.235	0.041
	EM	0.321	-0.087	0.116	0.238	0.056
	Regresyon	0.277	-0.105	-0.055	-0.061	0.021
%20 eksik	Listwise	0.259	-0.045	0.533	0.170	0.221
	Ortalama	0.434*	-0.131	0.319*	0.265	0.138
	EM	0.406*	-0.117	0.362*	0.230	0.162
	Regresyon	0.047	0.427	0.096	-0.185	0.121

\*  $p<0.05$       \*\*  $p<0.01$       \*\*\*  $p<0.001$



Tam veri seti ile %5 eksik veri setinin farklı atama yöntemleri ile tamamlanmasıyla elde edilen veri setlerinde yapılan regresyon analizlerinde ise Listwise atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_L^2=0.128$ ;  $R_T^2=0.142$ ). %10 eksik veri setlerinde yapılan regresyon analizlerinde Ortalama ve EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_O^2=0.132$ ;  $R_{EM}^2=0.172$ ;  $R_T^2=0.142$ ). %15 eksik veri setlerinde yapılan regresyon analizlerinde ise Listwise atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_L^2=0.102$ ;  $R_T^2=0.142$ ). %20 eksik setlerinde yapılan regresyon analizlerinde ise Ortalama ve EM atama yöntemlerinden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_O^2=0.138$ ;  $R_L^2=0.162$ ;  $R_T^2=0.142$ ).

Çizelge 4.6'da 50 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.6- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.218±1.131</b>	<b>0.536±1.237</b>	<b>0.95</b>	<b>48</b>	<b>0.35</b>
%5 eksik	Pairwise	0.218±1.131	0.426±0.272	-0.60	45	0.56
	Ortalama	0.218±1.131	0.413±1.190	-0.59	48	0.56
	EM	0.218±1.131	0.421±1.192	-0.62	48	0.54
	Regresyon	0.218±1.131	0.298±1.271	-0.24	48	0.81
%10 eksik	Pairwise	0.357±0.909	0.467±1.263	-0.34	44	0.74
	Ortalama	0.360±0.889	0.460±1.181	-0.34	48	0.74
	EM	0.349±0.890	0.467±1.183	-0.40	48	0.69
	Regresyon	0.316±0.913	0.459±1.246	-0.46	48	0.65
%15 eksik	Pairwise	0.159±1.163	0.444±0.290	-0.75	40	0.46
	Ortalama	0.170±1.114	0.406±1.119	-0.75	48	0.46
	EM	0.160±1.113	0.409±1.125	-0.78	48	0.44
	Regresyon	0.339±1.044	0.176±1.133	0.53	48	0.60
%20 eksik	Pairwise	0.191±1.224	0.522±1.317	-0.83	39	0.41
	Ortalama	0.255±1.091	0.496±1.204	-0.83	48	0.41
	EM	0.195±1.103	0.495±1.211	-0.91	48	0.37
	Regresyon	0.174±1.393	-0.030±1.349	0.53	48	0.60

Çizelge 4.7’de 50 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X2 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.7- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.401±1.122</b>	<b>0.093±0.862</b>	<b>1.09</b>	<b>48</b>	<b>0.28</b>
%5 eksik	Pairwise	0.389±1.129	0.093±0.862	1.02	45	0.32
	Ortalama	0.371±1.057	0.093±0.862	1.01	48	0.32
	EM	0.354±1.061	0.093±0.862	0.96	48	0.35
	Regresyon	0.293±1.093	0.093±0.862	0.72	48	0.48
%10 eksik	Pairwise	0.412±1.169	0.103±0.879	1.01	43	0.32
	Ortalama	0.386±1.069	0.108±0.861	1.01	48	0.32
	EM	0.379±1.078	0.115±0.863	0.96	48	0.34
	Regresyon	0.420±1.104	0.095±0.861	0.16	48	0.25
%15 eksik	Pairwise	0.551±1.109	0.168±0.864	1.25	40	0.22
	Ortalama	0.520±1.015	0.198±0.792	1.25	48	0.22
	EM	0.479±1.030	0.186±0.796	1.13	48	0.27
	Regresyon	0.405±0.818	0.181±0.815	0.97	48	0.34
%20 eksik	Pairwise	0.300±1.073	0.037±0.935	0.84	39	0.41
	Ortalama	0.280±0.981	0.064±0.834	0.84	48	0.41
	EM	0.300±0.983	0.110±0.850	0.73	48	0.47
	Regresyon	-0.062±0.916	0.154±1.083	-0.76	48	0.45

Çizelge 4.8’de 50 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X3 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir..

Çizelge 4.8- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.088±1.195</b>	<b>0.021±1.228</b>	<b>-0.32</b>	<b>48</b>	<b>0.75</b>
%5 eksik	Pairwise	-0.088±1.195	-0.098±1.142	0.03	45	0.98
	Ortalama	-0.088±1.095	-0.097±1.068	0.03	48	0.98
	EM	0.090±1.195	-0.100±1.082	0.03	48	0.98
	Regresyon	-0.090±1.195	-0.016±1.248	0.17	48	0.87
%10 eksik	Pairwise	-0.051±1.205	-0.027±1.140	-0.07	43	0.95
	Ortalama	-0.050±1.180	-0.029±1.040	-0.07	48	0.95
	EM	-0.027±1.186	-0.115±1.111	0.27	48	0.79
	Regresyon	-0.119±1.228	-0.247±1.205	0.37	48	0.71
%15 eksik	Pairwise	-0.130±1.167	0.103±1.251	-0.62	40	0.54
	Ortalama	-0.111±1.066	0.084±1.143	-0.63	48	0.54
	EM	-0.084±1.086	0.077±1.144	-0.51	48	0.61
	Regresyon	-0.352±1.081	-0.207±1.164	-0.46	48	0.65
%20 eksik	Pairwise	-0.203±1.133	0.032±1.239	-0.64	40	0.53
	Ortalama	-0.179±1.009	0.019±1.160	-0.64	48	0.52
	EM	-0.126±1.028	0.037±1.160	-0.53	48	0.60
	Regresyon	-0.257±0.117	0.332±0.968	1.99	48	0.052

Çizelge 4.9'da 50 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X4 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.9- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.107±0.926</b>	<b>-0.089±0.939</b>	<b>0.75</b>	<b>48</b>	<b>0.46</b>
%5 eksik	Pairwise	0.122±0.956	-0.149±0.909	1.00	45	0.32
	Ortalama	0.111±0.916	-0.144±0.890	1.00	48	0.32
	EM	0.069±0.938	-0.128±0.896	0.76	48	0.45
	Regresyon	0.068±0.944	-0.165±0.893	0.90	48	0.37
%10 eksik	Pairwise	0.088±0.946	-0.115±0.998	0.70	43	0.49
	Ortalama	0.080±0.906	-0.102±0.934	0.70	48	0.49
	EM	0.121±0.919	-0.071±0.955	0.72	48	0.47
	Regresyon	0.060±0.914	-0.088±0.975	0.55	48	0.58
%15 eksik	Pairwise	0.074±0.931	-0.213±0.909	1.00	40	0.32
	Ortalama	0.069±0.911	-0.167±0.769	0.99	48	0.33
	EM	0.084±0.912	-0.144±0.793	0.94	48	0.35
	Regresyon	0.121±0.766	0.026±0.901	0.40	48	0.69
%20 eksik	Pairwise	0.061±0.742	0.263±0.883	0.28	39	0.21
	Ortalama	0.036±0.680	-0.230±0.788	1.28	48	0.21
	EM	0.027±0.701	-0.209±0.797	1.11	48	0.27
	Regresyon	-0.065±0.636	0.056±0.724	-0.63	48	0.53

%5 eksik veri içeren veri setinin atama yöntemleri ile tamamlanmasından sonra elde edilen veri setlerindeki grup ortalamaları arasındaki farkın t testi ile araştırılması sonucunda, X1 ve X2 değişkenlerinde Pairwise atama yöntemi tam veri seti ile benzer sonuçlar vermiştir. X3 değişkeninde Regresyon atama yöntemi benzerlik göstermekte, X4 değişkeninde ise tam veri seti ile EM atama yöntemi sonuçları benzerlik göstermektedir. %10 ve %15 eksik veri setlerinde ortalamalar arasındaki farkların saptanmasında ise genel itibariyle EM atama yöntemi daha benzer sonuçlar vermiştir. %20 eksik veri setlerinde iki gruba ait ortalamalar arasındaki farkların saptanmasında genel itibariyle EM atama yöntemi daha benzer sonuçlar verse de, bir belirsizlik sözkonusudur. %20'lik veri setlerinde Regresyon atamada X2 değişkeni anlamlılık sınırında bulunmuştur. Bu değişkende tam veri setinde ortalamalar arasında fark bulunmazken ( $t=0.75$ ;  $p=0.46$ ), Regresyon atama yöntemi sonucunda elde edilen veri setinde ortalamalar arasındaki fark sınırda anlamlı bulunmuştur ( $t=1.99$ ;  $p=0.052$ ).

#### 4.1.2- Yüksek Korelasyonlu 50 Birimlik Veri Seti

50 birim içeren tam veri setine ilişkin değişkenler arasındaki korelasyonları gösteren Çizelge aşağıdadır. Çizelge 4.10'da görüleceği gibi değişkenler arasındaki korelasyon katsayıları oldukça yüksektir ( $r > 0.85$ ).

Çizelge 4.10- 50 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri

	X1	X2	X3	X4
X1	1	0.92 (P<0.001)	0.87 (P<0.001)	0.87 (P<0.001)
X2		1	0.92 (P<0.001)	0.90 (P<0.001)
X3			1	0.90 (P<0.001)
X4				1

Not: Çizelgede parantez içindeki değerler olasılık değerleridir.

Çizelge 4.11'de 50 birimden oluşan yüksek korelasyonlara sahip tam veri setinde sırasıyla %5, %10, %15 ve %20 değer eksiltilerek elde edilen veri setlerindeki eksik olma sayı ve yüzdeleri verilmiştir. Çizelge incelendiğinde eksiltmelerde verilen yüzde değerlerine yaklaşık azaltmalar gerçekleşmiştir. Bu farklılaşmalar tam rasgelelik koşullarının sağlanması için yerine koyma yöntemi ile eksiltme yönteminin kullanılmasından kaynaklanmaktadır. Ancak öngörülen yüzde eksiltme değerlerine yaklaşmıştır.

Çizelge 4.11- 50 birim içeren yüksek korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri

Değişken	Tam veri (n)	%5 Eksik		%10 Eksik		%15 Eksik		%20 Eksik	
		sayı	%	sayı	%	sayı	%	sayı	%
X1	50	3	6	5	10	8	16	9	18
X2	50	3	6	5	10	8	16	9	18
X3	50	3	6	5	10	8	16	9	18
X4	50	3	6	4	8	6	12	10	20
X5	50	0	0	0	0	0	0	0	0

Çizelge 4.12'de tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen ortalama ve standart sapma değerleri verilmiştir.

Yüksek korelasyonlu değişkenlere sahip 50 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde Regresyon ve EM atama yöntemleri, X1, X2 ve X3 değişkenleri göz önüne alındığında tam veri setinden elde edilen belirtici istatistiklere benzer sonuçlar vermiştir. Genel olarak incelendiğinde, %10'luk ve %15'lik eksik veri setinde, EM atama yöntemi, tam veri setinden elde edilen belirtici istatistiklere yakın sonuçlar vermiştir. Bu sonuçlar türetilen 4 değişkende de benzerlik göstermektedir. %20'lik eksik veri setinde, farklı atama yöntemleri ile tam veri setinden elde edilen belirtici istatistikler benzerlik göstermemektedir.

Çizelge 4.12- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri

		X1 $\bar{X} \pm S$	X2 $\bar{X} \pm S$	X3 $\bar{X} \pm S$	X4 $\bar{X} \pm S$
	<b>Tam Veri</b>	<b>-0.361±1.019</b>	<b>-0.356±1.018</b>	<b>-0.322±0.976</b>	<b>-0.353±1.061</b>
%5 eksik	Listwise	-0.185±1.013	-0.203±1.004	-0.165±0.954	-0.219±1.066
	Pairwise	-0.323±1.021	-0.331±1.007	-0.276±0.965	-0.357±1.059
	Ortalama	-0.323±0.989	-0.331±0.975	-0.276±0.935	-0.357±1.026
	EM	-0.353±1.012	-0.359±1.018	-0.305±0.957	-0.372±1.053
	Regresyon	-0.365±1.006	-0.360±1.004	-0.306±0.949	-0.385±1.043
%10 eksik	Listwise	-0.355±0.883	-0.319±0.928	-0.284±0.951	-0.327±1.001
	Pairwise	-0.361±0.936	-0.365±0.967	-0.355±1.012	-0.312±1.075
	Ortalama	-0.361±0.887	-0.365±0.916	-0.355±0.959	-0.312±1.031
	EM	-0.354±0.963	-0.364±0.973	-0.320±0.993	-0.362±1.071
	Regresyon	-0.336±0.974	-0.368±0.967	-0.324±0.983	-0.342±1.119
%15 eksik	Listwise	-0.304±1.046	-0.219±1.153	-0.245±1.107	-0.271±1.167
	Pairwise	-0.387±1.048	-0.313±1.049	-0.337±1.022	-0.319±1.073
	Ortalama	-0.387±0.959	-0.313±0.960	-0.337±0.935	-0.319±1.005
	EM	-0.384±1.041	-0.380±1.025	-0.332±0.982	-0.359±1.069
	Regresyon	-0.436±1.031	-0.397±1.019	-0.335±0.984	-0.362±1.080
%20 eksik	Listwise	-0.139±0.839	-0.249±0.896	-0.120±0.930	-0.336±1.022
	Pairwise	-0.346±0.997	-0.256±0.912	-0.313±0.975	-0.241±1.026
	Ortalama	-0.346±0.901	-0.256±0.824	-0.313±0.881	-0.241±0.916
	EM	-0.349±0.981	-0.361±0.969	-0.277±0.969	-0.392±1.075
	Regresyon	-0.310±1.055	-0.296±0.912	-0.310±0.989	-0.361±1.033

Çizelge 4.13'te tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkeni bağımlı ve X2, X3 ve X4 değişkenleri bağımsız değişkenler olarak alındığında, elde edilen regresyon katsayıları, bu katsayıların anlamlılık düzeyleri ve belirleyicilik katsayıları ( $R^2$ ) verilmiştir.

Yüksek korelasyonlu değişkenlere sahip 50 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde yapılan regresyon analizlerinde ise EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere

daha yakın çıkmıştır ( $R_{EM}^2=0.855$ ;  $R_T^2=0.847$ ). Regresyon atama yöntemiyle elde edilen veri setinin belirleyicilik katsayısı, olması gerekenden yüksek bulunmuştur ( $R_R^2=0.933$ ).

Çizelge 4.13- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları

		Sabit ( $b_0$ )	$b_1$	$b_2$	$b_3$	$R^2$
	<b>Tam Veri</b>	<b>-0.029</b>	<b>0.643***</b>	<b>0.100</b>	<b>0.201</b>	<b>0.847</b>
%5 eksik	Listwise	0.006	0.656**	0.010	0.257	0.817
	Ortalama	-0.020	0.517**	0.213	0.206	0.782
	EM	-0.018	0.681***	0.030	0.218	0.855
	Regresyon	-0.417***	0.541***	0.592***	-0.438***	0.933
%10 eksik	Listwise	-0.087	0.473*	0.159	0.220	0.765
	Ortalama	-0.080	0.466**	0.213	0.113	0.634
	EM	-0.042	0.387*	0.249	0.252	0.817
	Regresyon	0.111	0.500***	-0.074	0.541***	0.946
%15 eksik	Listwise	-0.091	0.444*	0.186	0.261	0.912
	Ortalama	-0.090	-0.046	0.759***	0.097	0.748
	EM	-0.038	0.423**	0.290	0.248	0.857
	Regresyon	0.042	0.581*	0.371	0.149	0.768
%20 eksik	Listwise	0.054	0.581*	0.131	0.095	0.732
	Ortalama	-0.070	0.059	0.793***	0.073	0.737
	EM	-0.030	0.463**	0.280	0.190	0.846
	Regresyon	0.074	0.536	0.316	0.284	0.722

\*  $p<0.05$       \*\*  $p<0.01$       \*\*\*  $p<0.001$

%10 eksik veri setlerinde yapılan regresyon analizlerinde ise EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2=0.817$ ;  $R_T^2=0.847$ ). Burada da %5'lik veri setlerinde olduğu gibi Regresyon atama yöntemiyle elde edilen veri setinin belirleyicilik katsayısı olması gerekenden yüksek bulunmuştur ( $R_R^2=0.946$ ). %15 eksik veri setlerinde yapılan regresyon analizlerinde ise EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2=0.857$ ;  $R_T^2=0.847$ ). Regresyon atama yöntemiyle elde edilen veri setinin belirleyicilik katsayısı olması gerekenden daha düşük bulunmuştur ( $R_R^2=0.768$ ). %20 eksik veri setlerinde ise EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2=0.846$ ;  $R_T^2=0.847$ ). Regresyon atama yöntemiyle elde edilen veri setinin belirleyicilik katsayısı olması gerekenden daha düşük bulunmuştur ( $R_R^2=0.722$ ).

Çizelge 4.14’de 50 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.14- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.505±1.091</b>	<b>-0.238±0.958</b>	<b>-0.92</b>	<b>48</b>	<b>0.36</b>
%5 eksik	Pairwise	-0.425±1.089	-0.240±0.977	-0.61	45	0.54
	Ortalama	-0.0416±1.039	-0.243±0.958	-0.61	48	0.54
	EM	-0.477±1.063	-0.248±0.958	-0.80	48	0.43
	Regresyon	-1.122±0.797	-1.249±0.735	0.58	48	0.56
%10 eksik	Pairwise	-0.640±0.994	-1.138±0.840	-1.84	43	0.07
	Ortalama	-0.603±0.929	-0.155±0.810	-1.83	48	0.07
	EM	-0.509±1.017	-0.221±0.885	-1.07	48	0.29
	Regresyon	-0.195±1.150	-0.552±1.293	1.02	48	0.31
%15 eksik	Pairwise	-0.496±1.209	-0.313±0.943	-0.55	40	0.59
	Ortalama	-0.467±1.032	-0.319±0.906	-0.54	48	0.59
	EM	-0.503±1.118	-0.281±0.938	-0.76	48	0.45
	Regresyon	-1.020±0.935	-0.528±1.129	-1.66	48	0.10
%20 eksik	Pairwise	-0.385±1.073	-0.319±0.963	-0.21	39	0.84
	Ortalama	-0.375±0.915	-0.322±0.906	-0.21	48	0.84
	EM	-0.420±0.996	-0.289±0.935	-0.48	48	0.64
	Regresyon	-0.317±1.031	-0.391±1.148	0.24	48	0.81

Çizelge 4.15’te 50 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X2 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.15- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.383±1.095</b>	<b>-0.333±0.968</b>	<b>-0.17</b>	<b>48</b>	<b>0.86</b>
%5 eksik	Pairwise	-0.414±1.111	-0.258±0.923	-0.53	45	0.60
	Ortalama	-0.410±1.085	-0.263±0.887	-0.53	48	0.60
	EM	-0.386±1.093	-0.336±0.959	-0.17	48	0.86
	Regresyon	-1.059±0.901	-1.126±0.880	-0.27	48	0.79
%10 eksik	Pairwise	-0.412±0.992	-0.333±0.968	-0.27	43	0.79
	Ortalama	-0.402±0.872	-0.333±0.968	-0.26	48	0.79
	EM	-0.402±0.977	-0.333±0.968	-0.25	48	0.81
	Regresyon	-0.298±1.018	-0.663±1.224	1.13	48	0.26
%15 eksik	Pairwise	-0.236±1.107	-0.376±1.020	0.43	40	0.67
	Ortalama	-0.250±1.002	-0.367±0.939	0.43	48	0.67
	EM	-0.381±1.093	-0.379±0.949	-0.01	48	0.99
	Regresyon	-0.839±0.769	-0.706±0.926	-0.54	48	0.59
%20 eksik	Pairwise	-0.408±1.016	-0.137±0.825	-0.94	39	0.35
	Ortalama	-0.375±0.895	0.155±0.761	-0.94	48	0.35
	EM	-0.438±1.013	-0.296±0.901	-0.52	48	0.60
	Regresyon	-0.396±0.862	-0.212±0.830	-0.77	48	0.45

Çizelge 4.16'da 50 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X3 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.16- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.353±1.025</b>	<b>-0.296±0.951</b>	<b>-0.20</b>	<b>48</b>	<b>0.84</b>
%5 eksik	Pairwise	-0.353±1.025	-0.201±0.918	-0.53	45	0.60
	Ortalama	-0.353±1.025	-0.210±0.864	-0.54	48	0.59
	EM	-0.353±1.025	-0.264±0.903	-0.33	48	0.75
	Regresyon	-1.160±0.934	-1.235±0.816	0.30	48	0.77
%10 eksik	Pairwise	-0.393±1.056	-0.321±0.995	-0.24	43	0.81
	Ortalama	-0.390±1.007	-0.325±0.936	-0.24	48	0.81
	EM	-0.357±1.019	-0.288±0.968	-0.25	48	0.81
	Regresyon	-0.281±0.975	-0.716±1.138	1.44	48	0.16
%15 eksik	Pairwise	-0.326±1.041	-0.348±1.027	0.07	40	0.94
	Ortalama	-0.326±1.017	-0.345±0.878	0.07	48	0.94
	EM	-0.365±1.034	-0.304±0.929	-0.22	48	0.83
	Regresyon	-0.731±0.781	-0.601±0.984	-0.51	48	0.61
%20 eksik	Pairwise	-0.329±1.009	-0.300±0.968	-0.09	39	0.93
	Ortalama	-0.326±0.912	-0.302±0.870	-0.10	48	0.93
	EM	-0.340±1.014	-0.224±0.912	-0.42	48	0.67
	Regresyon	-0.525±0.886	-0.518±0.867	-0.03	48	0.98

Çizelge 4.17'de 50 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X4 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.17- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.456±1.026</b>	<b>-0.265±0.876</b>	<b>-0.63</b>	<b>48</b>	<b>0.53</b>
%5 eksik	Pairwise	-0.483±1.279	-0.265±0.876	-0.69	45	0.49
	Ortalama	-0.466±1.190	-0.265±0.876	-0.69	48	0.49
	EM	-0.498±1.229	-0.265±0.876	-0.78	48	0.44
	Regresyon	-1.159±0.695	-1.249±0.575	0.50	48	0.62
%10 eksik	Pairwise	-0.419±1.298	-0.223±0.863	-0.61	44	0.55
	Ortalama	-0.409±1.238	-0.230±0.830	-0.61	48	0.54
	EM	-0.441±1.243	-0.295±0.898	-0.48	48	0.63
	Regresyon	-0.325±1.263	-0.715±1.423	1.02	48	0.32
%15 eksik	Pairwise	-0.516±1.329	-0.156±0.796	-1.11	42	0.27
	Ortalama	-0.490±1.237	-0.174±0.750	-1.11	48	0.27
	EM	-0.477±1.244	-0.258±0.883	-0.73	48	0.47
	Regresyon	-0.775±0.878	-0.646±1.067	-0.46	48	0.65
%20 eksik	Pairwise	-0.398±1.281	-0.113±0.767	-0.87	38	0.39
	Ortalama	-0.364±1.128	-0.137±0.691	-0.87	48	0.39
	EM	-0.492±0.198	-0.308±0.928	-0.61	48	0.54
	Regresyon	-0.381±0.975	-0.377±0.875	-0.02	48	0.99



Farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise %5'lik veri setlerinde genel itibariyle EM atama yöntemi, tam veri seti ile benzer sonuçlar verse de bir belirsizlik gözlenmektedir. %10 eksik veri setlerinde farklı gruplara ait ortalamalar arasındaki farkların saptanmasında ise genel itibariyle EM atama yöntemi benzer sonuçlar verse de %5'lik eksik veri setinde olduğu gibi bir belirsizlik söz konusudur. %15'lik veri setlerinde de EM atama yöntemi tam veri setinden elde edilen değerlere benzer sonuçlar vermiştir. %20 eksik veri setlerinde ise genel itibariyle EM atama yöntemi tam veri setinden elde edilen değerlere benzer sonuçlar verse de bir belirsizlik söz konusudur.

#### **4.2- 100 Birimlik Türetilmiş Veri Setleri**

Düşük ve yüksek korelasyonlara sahip 100 birimlik veri setlerine ilişkin analizler aşağıdaki bölümlerde verilmiştir.

##### **4.2.1- Düşük Korelasyonlu 100 Birimlik Veri Seti**

100 birim içeren tam veri setine ilişkin değişkenler arasındaki korelasyonları gösteren Çizelge aşağıdadır. Çizelge 4.18'de görüleceği gibi değişkenler arasındaki korelasyon katsayıları oldukça düşüktür ( $r < 0.20$ ).

Çizelge 4.18- 100 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri

	X1	X2	X3	X4
X1	1	0.01 (P>0.05)	0.06 (P>0.05)	0.11 (P>0.05)
X2		1	0.10 (P>0.05)	0.15 (P>0.05)
X3			1	0.11 (P>0.05)
X4				1

Not: Çizelgede parantez içindeki değerler olasılık değerleridir.

Çizelge 4.19'da 100 birimden oluşan düşük korelasyonlara sahip tam veri setinde sırasıyla %5, %10, %15 ve %20 değer eksiltilerek elde edilen veri setlerindeki eksik olma sayı ve yüzdeleri verilmiştir. Çizelge incelendiğinde eksiltmelerde verilen yüzde değerlerine yaklaşık azaltmalar gerçekleşmiştir. Bu farklılaşmalar tam rasgelelik koşullarının sağlanması için yerine koyma yöntemi ile eksiltme yönteminin

kullanılmasından kaynaklanmaktadır. Ancak öngörülen yüzde eksiltme değerlerine yaklaşılmıştır.

Çizelge 4.19- 100 birim içeren düşük korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri

Değişken	Tam veri (n)	%5 Eksik		%10 Eksik		%15 Eksik		%20 Eksik	
		sayı	%	sayı	%	sayı	%	sayı	%
X1	100	5	5	10	10	15	15	17	17
X2	100	5	5	9	9	13	13	19	19
X3	100	5	5	10	10	13	13	17	17
X4	100	5	5	10	10	13	13	19	19
X5	100	0	0	0	0	0	0	0	0

Çizelge 4.20'de tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen ortalama ve standart sapma değerleri verilmiştir.

Çizelge 4.20- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri

		X1 $\bar{X} \pm S$	X2 $\bar{X} \pm S$	X3 $\bar{X} \pm S$	X4 $\bar{X} \pm S$
<b>Tam Veri</b>		<b>-0.108±1.081</b>	<b>-0.057±1.024</b>	<b>-0.098±1.007</b>	<b>0.199±0.947</b>
%5 eksik	Listwise	-0.061±1.056	-0.065±1.015	-0.129±1.023	0.227±0.898
	Pairwise	-0.101±1.089	-0.050±1.026	-0.099±1.011	0.206±0.937
	Ortalama	-0.101±1.061	-0.050±1.000	-0.099±0.985	0.206±0.913
	EM	-0.106±1.089	-0.051±1.026	-0.102±1.011	0.207±0.937
	Regresyon	-0.133±1.081	-0.055±1.038	-0.087±1.002	0.163±0.951
%10 eksik	Listwise	-0.149±1.084	-0.057±0.946	-0.152±1.061	0.225±0.990
	Pairwise	-0.108±1.073	-0.100±1.025	-0.118±1.006	0.177±0.970
	Ortalama	-0.108±1.018	-0.100±0.977	-0.118±0.954	0.177±0.919
	EM	-0.112±1.073	-0.098±1.024	-0.117±1.006	0.174±0.970
	Regresyon	-0.163±1.075	-0.111±1.021	-0.142±0.980	0.171±0.956
%15 eksik	Listwise	-0.212±1.089	-0.050±0.997	0.039±0.942	0.232±0.943
	Pairwise	-0.086±1.070	-0.020±1.052	-0.107±0.999	0.211±0.922
	Ortalama	-0.086±0.985	-0.020±0.980	-0.107±0.932	0.211±0.859
	EM	-0.086±1.070	-0.022±1.051	-0.107±0.999	0.212±0.922
	Regresyon	-0.126±1.088	-0.004±1.028	-0.052±1.020	0.247±0.934
%20 eksik	Listwise	0.015±1.106	-0.056±0.905	-0.101±0.926	0.183±1.053
	Pairwise	-0.115±1.102	-0.031±1.057	-0.113±0.968	0.116±0.971
	Ortalama	-0.115±0.003	-0.031±0.950	-0.113±0.881	0.116±0.873
	EM	-0.124±1.103	-0.033±1.057	-0.117±0.968	0.117±0.971
	Regresyon	-0.106±1.104	-0.083±1.042	-0.093±0.927	0.030±1.004

Düşük korelasyonlu değişkenlere sahip 100 birimlik veri setlerinin analizleri sonucunda, %5'lik eksik veri setinde, EM atama yöntemi tam veri setinden elde edilen belirtici istatistiklere benzer sonuçlar vermiştir ( $P>0.05$ ). %10'luk eksik veri setinde,

aynı dağılım yapısına sahip 4 farklı değişken olmasına rağmen farklı sonuçlar elde edilmiştir ( $P < 0.05$ ). Genel olarak benzer ortalamalar Pairwise, Ortalama ve EM atama sonucunda oluşan veri setlerinden elde edilmiştir. %15'lik eksik veri setinde, EM, Ortalama ve Pairwise yöntemleri ile elde edilen ortalamalar, tam veri setinin ortalamalarına daha yakın bulunmuştur ( $P > 0.05$ ). %20'lik eksik veri setinde, Listwise ve Regresyon atama yöntemleri ile tam veri setinden elde edilen belirteci istatistikler arasında benzerlik gözlenmiştir ( $P > 0.05$ ).

Çizelge 4.21'de tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkeni bağımlı ve X2, X3 ve X4 değişkenleri bağımsız değişkenler olarak alındığında, elde edilen regresyon katsayıları, bu katsayıların anlamlılık düzeyleri ve belirleyicilik katsayıları ( $R^2$ ) verilmiştir.

Çizelge 4.21- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları

		Sabit ( $b_0$ )	$b_1$	$b_2$	$b_3$	$R^2$
	<b>Tam Veri</b>	<b>-0.129</b>	<b>-0.017</b>	<b>0.051</b>	<b>0.124</b>	<b>0.015</b>
%5 eksik	Listwise	-0.077	0.002	0.054	0.102	0.011
	Ortalama	-0.133	-0.030	0.068	0.179	0.029
	EM	-0.138	-0.028	0.069	0.183	0.030
	Regresyon	-0.554***	-0.006	-0.231	-0.343***	0.215
%10 eksik	Listwise	-0.166	-0.084	0.112	0.132	0.029
	Ortalama	-0.127	-0.040	0.057	0.117	0.014
	EM	-0.131	-0.042	0.060	0.124	0.016
	Regresyon	-0.110	0.239**	-0.802***	-0.232*	0.288
%15 eksik	Listwise	-0.242	-0.028	0.064	0.123	0.016
	Ortalama	-0.105	-0.020	-0.007	0.088	0.006
	EM	-0.108	-0.019	-0.008	0.098	0.007
	Regresyon	-0.171	-0.051	-0.168	0.088	0.031
%20 eksik	Listwise	0.013	-0.198	0.311	0.124	0.099
	Ortalama	-0.121	-0.080	0.118	0.146	0.032
	EM	-0.129	-0.085	0.144	0.164	0.043
	Regresyon	-0.209*	0.015	-0.151	0.342**	0.096

\*  $p < 0.05$       \*\*  $p < 0.01$       \*\*\*  $p < 0.001$

Düşük korelasyonlu değişkenlere sahip 100 birimlik veri setlerinin analizleri sonucunda, %5'lik eksik veri setinde, yapılan regresyon analizlerinde ise Listwise atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_L^2 = 0.011$ ;  $R_T^2 = 0.015$ ). Ayrıca Regresyon atama yönteminden elde edilen belirlilik katsayısı çok yüksek bulunmuştur ( $R_R^2 = 0.215$ ). %10'luk eksik veri setlerinde yapılan regresyon analizlerinde ise Ortalama ve EM

atama yöntemlerinden elde edilen belirleyicilik katsayıları, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2=0.016$ ;  $R_O^2=0.114$ ;  $R_T^2=0.015$ ). Regresyon atama yönteminden elde edilen belirlilik katsayısı oldukça yüksek bulunmuştur ( $R_R^2=0.288$ ). %15 eksik veri setlerinde yapılan regresyon analizlerinde ise Listwise atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_L^2=0.016$ ;  $R_T^2=0.015$ ). %20 eksik veri setlerinde yapılan regresyon analizlerinde ise Ortalama atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın bulunmuştur ( $R_O^2=0.032$ ;  $R_T^2=0.015$ ).

Çizelge 4.22’de 100 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.22- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.139±1.085</b>	<b>-0.072±1.087</b>	<b>-0.31</b>	<b>98</b>	<b>0.76</b>
%5 eksik	Pairwise	-0.136±1.102	-0.060±1.085	-0.34	93	0.74
	Ortalama	-0.134±1.081	-0.062±1.048	-0.34	98	0.74
	EM	-0.140±1.081	-0.066±1.049	-0.35	98	0.73
	Regresyon	-0.669±0.855	-0.662±0.916	-0.04	98	0.97
%10 eksik	Pairwise	-0.149±1.058	-0.068±1.099	-0.36	88	0.72
	Ortalama	-0.142±0.964	-0.069±1.087	-0.36	98	0.72
	EM	-0.150±0.964	-0.068±1.087	-0.40	98	0.69
	Regresyon	-0.181±1.043	-0.097±0.778	-0.45	98	0.65
%15 eksik	Pairwise	-0.031±1.089	-0.150±1.057	0.51	83	0.62
	Ortalama	-0.039±1.003	-0.140±0.972	0.51	98	0.61
	EM	-0.041±1.003	-0.138±0.973	0.49	98	0.63
	Regresyon	-0.166±1.141	-0.092±1.188	-0.32	98	0.75
%20 eksik	Pairwise	-0.154±1.108	-0.073±1.109	-0.33	81	0.74
	Ortalama	-0.146±0.986	-0.079±1.032	-0.34	98	0.74
	EM	-0.149±0.988	-0.095±1.035	-0.27	98	0.79
	Regresyon	-0.111±1.038	-0.225±0.951	0.57	98	0.57

Çizelge 4.23’te 100 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X2 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.23- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.026±1.165</b>	<b>-0.155±0.829</b>	<b>0.88</b>	<b>98</b>	<b>0.38</b>
%5 eksik	Pairwise	0.045±1.160	-0.160±0.846	0.97	93	0.34
	Ortalama	0.039±1.127	-0.155±0.827	0.97	98	0.34
	EM	0.038±1.127	-0.155±0.828	0.96	98	0.34
	Regresyon	0.138±1.394	0.321±1.416	-0.65	98	0.52
%10 eksik	Pairwise	-0.024±1.182	-0.186±0.819	0.75	89	0.45
	Ortalama	-0.032±1.114	-0.180±0.792	0.75	98	0.45
	EM	-0.032±1.114	-0.175±0.793	0.73	98	0.47
	Regresyon	0.076±1.182	-0.103±1.372	0.70	98	0.48
%15 eksik	Pairwise	0.037±1.198	-0.096±0.823	0.58	85	0.56
	Ortalama	0.033±1.152	-0.081±0.737	0.58	98	0.56
	EM	0.030±1.152	-0.083±0.738	0.57	98	0.57
	Regresyon	0.058±1.115	-0.060±1.133	0.52	98	0.60
%20 eksik	Pairwise	0.127±1.229	-0.210±0.798	1.44	79	0.15
	Ortalama	0.095±1.096	-0.179±0.727	1.44	98	0.15
	EM	0.093±1.097	-0.180±0.729	1.44	98	0.15
	Regresyon	0.008±1.330	-0.196±1.272	0.78	98	0.44

Çizelge 4.24'de 100 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X3 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.24- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.152±1.023</b>	<b>-0.036±0.995</b>	<b>-0.57</b>	<b>98</b>	<b>0.57</b>
%5 eksik	Pairwise	-0.163±1.030	-0.020±0.092	-0.68	93	0.50
	Ortalama	-0.162±1.020	-0.027±0.947	-0.68	98	0.50
	EM	-0.159±1.020	-0.035±0.949	-0.63	98	0.53
	Regresyon	0.146±0.710	0.104±0.675	0.30	98	0.76
%10 eksik	Pairwise	-0.176±0.997	-0.049±1.025	-0.59	88	0.56
	Ortalama	-0.170±0.949	-0.056±0.967	-0.59	98	0.56
	EM	-0.166±0.949	-0.059±0.967	-0.56	98	0.58
	Regresyon	0.081±0.722	-0.040±0.711	0.85	98	0.40
%15 eksik	Pairwise	-0.137±0.999	-0.075±1.011	-0.29	85	0.78
	Ortalama	-0.132±0.900	-0.077±0.978	-0.29	98	0.77
	EM	-0.132±0.900	-0.077±0.977	-0.30	98	0.77
	Regresyon	-0.021±0.963	-0.274±1.127	1.21	98	0.23
%20 eksik	Pairwise	-0.120±0.970	-0.104±0.978	-0.07	81	0.94
	Ortalama	-0.119±0.874	-0.106±0.899	-0.07	98	0.94
	EM	-0.113±0.875	-0.122±0.901	0.05	98	0.96
	Regresyon	-0.128±0.884	-0.242±1.025	0.60	98	0.55

Çizelge 4.25'de 100 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X4 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.25- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.104±0.877</b>	<b>0.310±1.023</b>	<b>-1.08</b>	<b>98</b>	<b>0.28</b>
%5 eksik	Pairwise	0.093±0.872	0.337±1.001	-1.27	93	0.21
	Ortalama	0.099±0.847	0.331±0.979	-1.27	98	0.21
	EM	-0.101±0.848	0.331±0.979	-1.26	98	0.21
	Regresyon	-0.222±1.007	-0.257±0.935	0.18	98	0.86
%10 eksik	Pairwise	-0.101±0.904	0.268±1.046	-0.81	88	0.42
	Ortalama	0.108±0.861	0.258±0.987	-0.81	98	0.42
	EM	0.107±0.861	0.254±0.988	-0.80	98	0.43
	Regresyon	0.046±0.938	0.046±0.768	-0.01	98	0.99
%15 eksik	Pairwise	0.134±0.839	0.301±1.014	-0.84	85	0.40
	Ortalama	0.144±0.782	0.289±0.944	-0.84	98	0.40
	EM	0.145±0.782	0.291±0.944	-0.84	98	0.40
	Regresyon	0.075±0.844	0.297±0.957	-1.23	98	0.22
%20 eksik	Pairwise	-0.010±0.851	0.252±1.080	-1.22	79	0.23
	Ortalama	0.018±0.751	0.231±0.993	-1.22	98	0.23
	EM	0.016±0.753	0.237±0.994	-1.27	98	0.21
	Regresyon	0.156±0.856	-0.058±0.912	1.21	98	0.23

Düşük korelasyonlu değişkenlere sahip 100 birimlik veri setlerinin analizleri sonucunda, %5'lik eksik veri setinde, farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise genel itibariyle Pairwise, Ortalama ve EM atama yöntemleri birbirlerine benzer sonuçlar vermiştir. %10 eksik veri setlerinde farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise genel itibariyle Pairwise ve Ortalama atama yöntemleri tam veri setine benzer sonuçlar vermiştir. %15 eksik veri setlerinde ise genel itibariyle Regresyon ve EM atama yöntemleri tam veri setine yakın sonuçlar vermesine rağmen ortalamalar arasında fark bulunmaktadır. Son olarak %20 eksik veri setlerinde genel itibariyle Regresyon atama yöntemi tam veri seti ile daha benzer sonuçlar verse de bir belirsizlik sözkonusudur.

#### **4.2.2- Yüksek Korelasyonlu 100 Birimlik Veri Seti**

100 birim içeren tam veri setine ilişkin değişkenler arasındaki korelasyonları gösteren Çizelge aşağıdadır. Çizelge 4.26'da görüleceği gibi değişkenler arasındaki korelasyon katsayıları oldukça yüksektir ( $r > 0.85$ ).

Çizelge 4.26- 100 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri

	X1	X2	X3	X4
X1	1	0.90 (P<0.001)	0.90 (P<0.001)	0.89 (P<0.001)
X2		1	0.91 (P<0.001)	0.90 (P<0.001)
X3			1	0.89 (P<0.001)
X4				1

Not: Çizelgede parantez içindeki değerler olasılık değerleridir.

Çizelge 4.27’de 100 birimden oluşan düşük korelasyonlara sahip tam veri setinde sırasıyla %5, %10, %15 ve %20 değer eksiltilerek elde edilen veri setlerindeki eksik olma sayı ve yüzdeleri verilmiştir. Çizelge incelendiğinde eksiltmelerde verilen yüzde değerlerine yaklaşık azaltmalar gerçekleşmiştir. Bu farklılaşmalar tam rasgelelik koşullarının sağlanması için yerine koyma yöntemi ile eksiltme yönteminin kullanılmasından kaynaklanmaktadır. Ancak öngörülen yüzde eksiltme değerlerine yaklaşılmıştır.

Çizelge 4.27- 100 birim içeren yüksek korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri

Değişken	Tam veri (n)	%5 Eksik		%10 Eksik		%15 Eksik		%20 Eksik	
		sayı	%	sayı	%	sayı	%	sayı	%
X1	100	5	5	9	9	14	14	19	19
X2	100	5	5	10	10	14	14	18	18
X3	100	5	5	8	8	13	13	17	17
X4	100	5	5	10	10	14	14	18	18
X5	100	0	0	0	0	0	0	0	0

Çizelge 4.28’de tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen ortalama ve standart sapma değerleri verilmiştir.

Yüksek korelasyonlu değişkenlere sahip 100 birimlik veri setlerinin analizleri sonucunda %5’lik ve %10’luk eksik veri setinde, EM atama yöntemi, tam veri setinden elde edilen belirtici istatistiklere oldukça yakın sonuçlar vermiştir (P>0.05). %15’lik eksik veri setinde, Listwise ve Regresyon atama yöntemleri aynı sonuçları vermiştir. Fakat EM atama yöntemi ile elde edilen ortalamalar tam veri setinin ortalamalarına daha yakın bulunmuştur (P>0.05). Fakat bu ortalamalar yinede birbirinden oldukça farklıdır. %20’lik eksik veri setinde ise, sadece EM atama yöntemi ile tam veri setinden elde edilen belirtici istatistikler benzerlik göstermektedir (P>0.05).

Çizelge 4.28- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri

		X1 ( $\bar{X} \pm S$ )	X2 ( $\bar{X} \pm S$ )	X3 ( $\bar{X} \pm S$ )	X4 ( $\bar{X} \pm S$ )
	<b>Tam Veri</b>	<b>0.074±0.968</b>	<b>0.120±0.969</b>	<b>0.146±0.991</b>	<b>0.150±0.979</b>
%5 eksik	Listwise	0.035±0.999	0.087±0.998	0.098±0.995	0.111±1.009
	Pairwise	0.057±0.974	0.094±0.974	0.179±0.990	0.148±0.991
	Ortalama	0.057±0.949	0.094±0.950	0.179±0.968	0.148±0.966
	EM	0.085±0.975	0.116±0.964	0.154±0.986	0.154±0.985
	Regresyon	0.035±0.999	0.087±0.998	0.098±0.995	0.111±1.009
%10 eksik	Listwise	0.012±1.003	0.032±0.959	0.043±1.035	0.102±1.027
	Pairwise	0.045±0.999	0.131±0.950	0.142±1.009	0.103±0.995
	Ortalama	0.045±0.953	0.131±0.901	0.142±0.168	0.103±0.943
	EM	0.071±0.975	0.122±0.944	0.128±0.996	0.158±1.006
	Regresyon	0.012±1.003	0.032±0.959	0.043±1.035	0.102±1.027
%15 eksik	Listwise	0.105±1.015	0.108±1.003	0.146±1.005	0.150±1.040
	Pairwise	0.090±0.999	0.105±0.973	0.166±1.019	0.171±0.975
	Ortalama	0.090±0.926	0.105±0.902	0.166±0.950	0.171±0.903
	EM	0.077±0.973	0.110±0.988	0.154±0.986	0.139±0.994
	Regresyon	0.105±1.015	0.108±1.003	0.146±1.005	0.150±1.040
%20 eksik	Listwise	0.035±0.969	0.014±0.980	0.035±0.999	0.017±1.008
	Pairwise	-0.016±0.926	0.143±1.011	0.157±1.007	0.192±0.994
	Ortalama	-0.16±0.832	0.143±0.915	0.157±0.916	0.192±0.899
	EM	0.097±0.964	0.121±0.985	0.151±1.003	0.156±0.982
	Regresyon	0.035±0.969	0.014±0.980	0.035±0.999	0.017±1.008

Çizelge 4.29'da tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkeni bağımlı ve X2, X3 ve X4 değişkenleri bağımsız değişkenler olarak alındığında, elde edilen regresyon katsayıları, bu katsayıların anlamlılık düzeyleri ve belirleyicilik katsayıları ( $R^2$ ) verilmiştir.

Çizelge 4.29- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları

		Sabit ( $b_0$ )	$b_1$	$b_2$	$b_3$	$R^2$
	<b>Tam Veri</b>	<b>-0.058</b>	<b>0.280*</b>	<b>0.361***</b>	<b>0.307**</b>	<b>0.861</b>
%5 eksik	Listwise	-0.061	0.267*	0.425***	0.274**	0.874
	Ortalama	-0.070	0.370**	0.241*	0.323**	0.814
	EM	-0.052	0.254*	0.378***	0.329**	0.875
	Regresyon	-0.064	0.301**	0.327**	0.322**	0.861
	Listwise	-0.049	0.264*	0.317**	0.389***	0.890
%10 eksik	Ortalama	-0.080	0.277**	0.419***	0.275**	0.814
	EM	-0.06	0.208	0.342**	0.402***	0.879
	Regresyon	-0.06	0.196	0.413***	0.345***	0.870
	Listwise	-0.03	0.125	0.489**	0.343**	0.863
	Ortalama	-0.050	0.366***	0.367***	0.213*	0.756
%15 eksik	EM	-0.05	0.211	0.375***	0.359***	0.875
	Regresyon	-0.07	0.364***	0.378***	0.226*	0.874
	Listwise	0.015	0.413*	0.263	0.268*	0.883
	Ortalama	-0.134*	0.338**	0.302**	0.115	0.597
	EM	-0.03	0.359**	0.402***	0.167	0.882
%20 eksik	Regresyon	-0.04	0.362**	0.436**	0.085	0.838

\*  $p < 0.05$       \*\*  $p < 0.01$       \*\*\*  $p < 0.001$



Yüksek korelasyonlu değişkenlere sahip 100 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde, yapılan regresyon analizlerinde ise Regresyon atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere oldukça yakın çıkmıştır ( $R_R^2=0.861$ ;  $R_T^2=0.861$ ). Fakat çizelge 5.29 incelendiğinde, EM atama yöntemi ile elde edilen regresyon katsayıları ile tam veri setinden elde edilen regresyon katsayıları ve anlamlılıklarının birbirlerine daha benzer oldukları görülmektedir ( $R_{EM}^2=0.875$ ). %10'luk eksik veri setlerinde yapılan regresyon analizlerinde ise Regresyon atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_R^2=0.870$ ;  $R_T^2=0.861$ ). Listwise ve EM atama yöntemleriyle elde edilen veri setlerinin belirleyicilik katsayıları olması gerekenden yüksek bulunmuştur ( $R_L^2=0.890$  ;  $R_{EM}^2=0.879$ ). %15'lik eksik veri setinde yapılan regresyon analizlerinde ise Listwise atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_L^2=0.863$ ;  $R_T^2=0.861$ ). %20'lik eksik veri setlerinde yapılan regresyon analizlerinde ise EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2=0.882$ ;  $R_T^2=0.861$ ). Regresyon atama yöntemiyle elde edilen veri setinin belirleyicilik katsayısı olması gerekenden daha düşük bulunmuştur ( $R_R^2=0.838$ ).

Çizelge 4.30'da 100 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.30- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.087±1.015</b>	<b>0.060±0.927</b>	<b>0.14</b>	<b>98</b>	<b>0.89</b>
%5 eksik	Pairwise	0.087±1.015	0.021±0.935	0.33	93	0.74
	Ortalama	0.087±1.015	0.025±0.885	0.33	98	0.74
	EM	0.087±1.015	0.083±0.934	0.02	98	0.98
	Regresyon	0.087±1.015	0.047±0.919	0.21	98	0.84
%10 eksik	Pairwise	0.067±1.051	0.023±0.954	0.21	89	0.84
	Ortalama	0.065±0.998	0.025±0.914	0.21	98	0.84
	EM	0.090±1.020	0.051±0.922	0.20	98	0.84
	Regresyon	0.085±1.023	0.057±0.928	0.15	98	0.88
%15 eksik	Pairwise	0.075±1.051	0.106±0.953	-0.14	84	0.89
	Ortalama	0.077±0.986	0.103±0.970	0.14	98	0.89
	EM	0.080±1.015	0.073±0.913	0.04	98	0.97
	Regresyon	0.055±1.023	0.056±0.950	-0.01	98	0.99
%20 eksik	Pairwise	0.052±0.899	-0.089±0.959	0.68	79	0.50
	Ortalama	0.040±0.815	-0.074±0.854	0.68	98	0.50
	EM	0.110±0.955	0.083±0.952	0.14	98	0.89
	Regresyon	0.094±0.935	0.073±0.962	0.11	98	0.91

Çizelge 4.31'de 100 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X2 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.31- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.153±1.019</b>	<b>0.085±0.924</b>	<b>0.35</b>	<b>98</b>	<b>0.73</b>
%5 eksik	Pairwise	0.098±1.025	0.090±0.933	0.04	93	0.97
	Ortalama	0.097±0.983	0.090±0.923	0.04	98	0.97
	EM	0.143±1.004	0.087±0.923	0.29	98	0.77
	Regresyon	0.140±0.999	0.076±0.928	0.33	98	0.74
%10 eksik	Pairwise	0.244±0.986	0.022±0.913	1.11	88	0.27
	Ortalama	0.228±0.915	0.029±0.884	1.11	98	0.57
	EM	0.170±0.970	0.073±0.912	0.51	98	0.61
	Regresyon	0.182±0.970	0.081±0.928	0.53	98	0.60
%15 eksik	Pairwise	0.115±1.038	0.096±0.916	0.09	84	0.93
	Ortalama	0.113±0.951	0.097±0.857	0.09	98	0.93
	EM	0.141±1.019	0.077±0.944	0.32	98	0.75
	Regresyon	0.148±1.008	0.084±0.934	0.33	98	0.74
%20 eksik	Pairwise	0.219±1.093	0.067±0.930	0.68	80	0.50
	Ortalama	0.204±0.978	0.079±0.850	0.68	98	0.50
	EM	0.149±1.042	0.092±0.908	0.29	98	0.77
	Regresyon	0.129±1.062	0.099±0.948	0.71	98	0.88

Çizelge 4.32'de 100 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X3 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.32- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.125±0.983</b>	<b>0.167±1.009</b>	<b>-0.21</b>	<b>98</b>	<b>0.84</b>
%5 eksik	Pairwise	0.198±0.966	0.159±1.024	0.19	93	0.85
	Ortalama	0.197±0.937	0.159±1.002	0.19	98	0.85
	EM	0.148±0.967	0.161±1.008	-0.07	98	0.95
	Regresyon	0.157±0.953	0.172±1.020	-0.08	98	0.94
%10 eksik	Pairwise	0.147±0.993	0.136±1.040	0.05	90	0.96
	Ortalama	0.147±0.973	0.137±0.973	0.05	98	0.96
	EM	0.122±0.982	0.135±1.009	-0.07	98	0.95
	Regresyon	0.119±0.985	0.115±1.012	0.02	98	0.98
%15 eksik	Pairwise	0.163±1.029	0.170±1.021	-0.03	85	0.98
	Ortalama	0.164±0.954	0.169±0.956	-0.03	98	0.98
	EM	0.136±0.983	0.174±0.982	-0.19	98	0.85
	Regresyon	0.099±1.015	0.168±0.993	-0.34	98	0.73
%20 eksik	Pairwise	0.174±1.002	0.141±1.023	0.15	81	0.88
	Ortalama	0.171±0.885	0.143±0.957	0.15	98	0.88
	EM	0.134±0.992	0.170±1.003	-0.18	98	0.86
	Regresyon	0.146±0.986	0.175±1.015	-0.15	98	0.89

Çizelge 4.33'te 100 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X4 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.33- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.104±1.056</b>	<b>0.198±0.899</b>	<b>-0.48</b>	<b>98</b>	<b>0.63</b>
%5 eksik	Pairwise	0.115±1.065	0.183±0.917	-0.33	93	0.74
	Ortalama	0.117±1.044	1.181±0.886	-0.33	98	0.74
	EM	0.114±1.068	0.196±0.892	-0.41	98	0.68
	Regresyon	0.098±1.077	0.192±0.890	-0.47	98	0.64
%10 eksik	Pairwise	0.052±1.080	0.155±0.906	-0.49	88	0.63
	Ortalama	0.057±1.025	0.150±0.858	-0.49	98	0.63
	EM	0.121±1.076	0.195±0.922	-0.37	98	0.71
	Regresyon	0.130±1.083	0.179±0.899	-0.25	98	0.81
%15 eksik	Pairwise	0.155±1.053	0.188±0.899	-0.16	84	0.88
	Ortalama	0.157±0.976	0.185±0.831	-0.16	98	0.88
	EM	0.114±1.060	0.165±0.906	-0.26	98	0.80
	Regresyon	0.110±1.065	0.182±0.893	-0.37	98	0.71
%20 eksik	Pairwise	0.176±1.064	0.208±0.935	-0.15	80	0.88
	Ortalama	0.179±0.940	0.206±0.864	-0.15	98	0.88
	EM	0.123±1.018	0.191±0.922	-0.35	98	0.73
	Regresyon	0.096±1.025	0.181±0.944	-0.43	98	0.67

Farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise %5 eksik veri setlerinde genel itibariyle Regresyon atama yöntemi, tam veri seti ile benzer sonuçlar vermiştir ( $P>0.05$ ). EM, Ortalama ve Pairwise atama yöntemleri ise benzerlik bakımından Regresyon analizinden sonra gelmektedir ( $P>0.05$ ). %10'luk eksik veri setlerinde genel itibariyle EM atama yöntemi, %15'lik eksik veri setlerinde de Regresyon atama yöntemi tam veri setine benzer sonuçlar vermesine rağmen bu veri setlerinden elde edilen sonuçlarda belirsizlik sözkonusudur. %20'lik eksik veri setlerinde ise genel itibariyle EM atama yöntemi tam veri setinden elde edilen değerlere benzer sonuçlar vermiştir ( $P>0.05$ ).

#### **4.3- 200 Birimlik Türetilmiş Veri Setleri**

Düşük ve yüksek korelasyonlara sahip 200 birimlik veri setlerine ilişkin analizler aşağıdaki bölümlerde verilmiştir.

#### 4.3.1- Düşük Korelasyonlu 200 Birimlik Veri Seti

200 birim içeren tam veri setine ilişkin değişkenler arasındaki korelasyonları gösteren Çizelge aşağıdadır. Çizelge 4.34'te görüleceği gibi değişkenler arasındaki korelasyon katsayıları oldukça düşüktür ( $r < 0.20$ ).

Çizelge 4.34- 100 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri

	X1	X2	X3	X4
X1	1	0.09 (P>0.05)	0.16 (P<0.05)	0.04 (P>0.05)
X2		1	0.15 (P<0.05)	-0.08(P>0.05)
X3			1	0.11 (P>0.05)
X4				1

Not: Çizelgede parantez içindeki değerler olasılık değerleridir.

Çizelge 4.35'te 200 birimden oluşan düşük korelasyonlara sahip tam veri setinde sırasıyla %5, %10, %15 ve %20 değer eksiltilerek elde edilen veri setlerindeki eksik olma sayı ve yüzdeleri verilmiştir. Çizelge incelendiğinde eksiltmelerde verilen yüzde değerlerine yaklaşık azaltmalar gerçekleşmiştir. Bu farklılaşmalar tam rasgelelik koşullarının sağlanması için yerine koyma yöntemi ile eksiltme yönteminin kullanılmasından kaynaklanmaktadır. Ancak öngörülen yüzde eksiltme değerlerine yaklaşılmıştır.

Çizelge 4.35- 200 birim içeren düşük korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri

Değişken	Tam veri (n)	%5 Eksik		%10 Eksik		%15 Eksik		%20 Eksik	
		sayı	%	sayı	%	sayı	%	sayı	%
X1	200	10	5	20	10	27	13.5	33	16.5
X2	200	10	5	19	9.5	28	14	37	18.5
X3	200	10	5	20	10	27	13.5	38	19
X4	200	10	5	18	9	30	15	34	17
X5	200	0	0	0	0	0	0	0	0

Çizelge 4.36'da tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen ortalama ve standart sapma değerleri verilmiştir.

Çizelge 4.36- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri

		X1 $\bar{X} \pm S$	X2 $\bar{X} \pm S$	X3 $\bar{X} \pm S$	X4 $\bar{X} \pm S$
	<b>Tam Veri</b>	<b>0.100±1.027</b>	<b>0.024±0.951</b>	<b>0.068±1.026</b>	<b>-0.020±0.998</b>
%5 eksik	Listwise	0.070±1.022	-0.012±0.958	0.067±1.033	-0.033±1.013
	Pairwise	0.095±1.039	0.008±0.951	0.076±1.023	-0.010±1.011
	Ortalama	0.095±1.013	0.008±0.926	0.076±0.998	-0.010±0.985
	EM	0.097±1.039	0.005±0.950	0.083±1.024	-0.011±1.010
	Regresyon	0.070±1.022	-0.012±0.958	0.067±1.033	-0.033±1.013
%10 eksik	Listwise	0.203±1.041	0.080±0.970	0.112±1.017	-0.006±1.081
	Pairwise	0.102±1.018	0.068±0.962	0.045±1.022	-0.032±1.017
	Ortalama	0.102±0.965	0.068±0.914	0.045±0.969	-0.032±0.970
	EM	0.093±1.017	0.065±0.962	0.032±1.021	-0.031±1.017
	Regresyon	0.203±1.041	0.080±0.970	0.112±1.017	-0.006±1.081
%15 eksik	Listwise	0.004±0.957	0.022±0.859	0.142±1.094	0.081±1.079
	Pairwise	0.079±1.029	0.025±0.940	0.073±1.040	-0.015±1.032
	Ortalama	0.079±0.957	0.025±0.871	0.073±0.967	-0.015±0.951
	EM	0.078±1.031	0.031±0.941	0.073±1.040	-0.017±1.032
	Regresyon	0.004±0.957	0.022±0.859	0.142±1.094	0.081±1.079
%20 eksik	Listwise	0.048±0.975	-0.012±0.932	-0.019±1.005	-0.026±0.912
	Pairwise	0.167±1.036	0.048±0.960	0.007±1.035	-0.028±1.015
	Ortalama	0.167±0.946	0.048±0.866	0.007±0.931	-0.028±0.924
	EM	0.169±1.038	0.051±0.961	0.005±1.036	-0.025±1.016
	Regresyon	0.048±0.975	-0.012±0.932	-0.019±1.005	-0.026±0.912

Düşük korelasyonlu değişkenlere sahip 200 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde, Listwise ve Regresyon atama yöntemleri kendi aralarında oldukça benzer sonuçları vermiştir ( $P>0.05$ ). Fakat burada EM atama yöntemi tam veri setinden elde edilen belirtici istatistiklere daha yakın sonuçlar vermiştir ( $P>0.05$ ). %10'luk eksik veri setlerinde, Listwise ve Regresyon atama yöntemleri ile Pairwise ve Ortalama atama yöntemleri kendi aralarında hemen hemen aynı sonuçları vermiştir ( $P>0.05$ ). Fakat burada da %5'lik veri setinde olduğu gibi EM atama yöntemi tam veri setinden elde edilen belirtici istatistiklere daha yakın sonuçlar ortaya koymuştur. %15'lik eksik veri setlerinde EM, Ortalama ve Pairwise atama yöntemlerinden elde edilen ortalamalar tam veri setinin ortalamalarına daha yakın bulunmuştur ( $P>0.05$ ). %20'lik eksik veri setlerinde, Pairwise ve Ortalama atama yöntemleri ile tam veri setinden elde edilen belirtici istatistikler arasında benzerlik gözlenmiştir ( $P>0.05$ ).

Çizelge 4.37'de tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkeni bağımlı ve X2, X3 ve X4 değişkenleri bağımsız

değişkenler olarak alındığında, elde edilen regresyon katsayıları, bu katsayıların anlamlılık düzeyleri ve belirleyicilik katsayıları ( $R^2$ ) verilmiştir.

Çizelge 4.37- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları

		Sabit ( $b_0$ )	$b_1$	$b_2$	$b_3$	$R^2$
	<b>Tam Veri</b>	<b>0.089</b>	<b>0.072</b>	<b>0.146*</b>	<b>0.026</b>	<b>0.030</b>
%5 eksik	Listwise	0.062	0.053	0.128	-0.015	0.022
	Ortalama	0.085	0.082	0.125	0.021	0.025
	EM	0.085	0.086	0.135	0.024	0.030
	Regresyon	0.068	0.098	0.089	0.046	0.021
%10 eksik	Listwise	0.173	0.096	0.198*	-0.033	0.049
	Ortalama	0.090	0.052	0.181*	-0.010	0.037
	EM	0.082	0.055	0.218**	-0.016	0.053
	Regresyon	0.092	0.001	0.163*	-0.026	0.026
%15 eksik	Listwise	-0.021	0.111	0.112	0.077	0.038
	Ortalama	0.068	0.098	0.123	0.052	0.030
	EM	0.063	0.200*	0.144*	0.071	0.069
	Regresyon	0.049	0.187*	0.171*	0.027	0.072
%20 eksik	Listwise	0.056	0.136	0.116	0.160	0.066
	Ortalama	0.165*	0.068	0.084	0.093	0.022
	EM	0.167*	0.093	0.141	0.147*	0.057
	Regresyon	0.159*	0.044	0.071	0.128	0.024

\*  $p < 0.05$       \*\*  $p < 0.01$       \*\*\*  $p < 0.001$

Düşük korelasyonlu değişkenlere sahip 200 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde yapılan regresyon analizlerinde ise EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere oldukça yakın çıkmıştır ( $R_{EM}^2 = 0.030$ ;  $R_T^2 = 0.030$ ). %10'luk eksik veri setlerinde yapılan regresyon analizlerinde Regresyon atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_R^2 = 0.026$ ;  $R_T^2 = 0.030$ ). %15'lik eksik veri setlerinde yapılan regresyon analizlerinde Ortalama atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_O^2 = 0.030$ ;  $R_T^2 = 0.030$ ). %20'lik eksik veri setinde yapılan regresyon analizlerinde ise Ortalama ve Regresyon atama yöntemlerinden elde edilen belirleyicilik katsayıları, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_R^2 = 0.024$ ;  $R_O^2 = 0.022$ ;  $R_T^2 = 0.030$ ).

Çizelge 4.38'de 200 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir..

Çizelge 4.38- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.065±1.086</b>	<b>0.131±0.976</b>	<b>-0.45</b>	<b>198</b>	<b>0.65</b>
%5 eksik	Pairwise	0.047±1.101	0.139±0.983	-0.61	188	0.54
	Ortalama	0.049±1.077	0.137±0.955	-0.61	198	0.54
	EM	0.049±1.078	0.139±0.955	-0.63	198	0.53
	Regresyon	-0.003±1.107	0.140±0.983	-0.97	198	0.33
%10 eksik	Pairwise	0.069±1.085	0.131±0.961	-0.41	178	0.69
	Ortalama	0.073±1.019	0.129±0.919	-0.41	198	0.69
	EM	0.068±1.020	0.115±0.922	-0.34	198	0.73
	Regresyon	0.038±1.060	0.150±1.018	-0.77	198	0.45
%15 eksik	Pairwise	0.062±1.090	0.094±0.974	-0.20	171	0.84
	Ortalama	0.064±1.030	0.092±0.892	-0.21	198	0.84
	EM	0.066±1.033	0.089±0.899	-0.16	198	0.87
	Regresyon	0.094±1.085	0.057±0.958	0.26	198	0.80
%20 eksik	Pairwise	0.113±1.071	0.217±1.005	-0.65	165	0.52
	Ortalama	0.120±0.994	0.208±0.904	-0.65	198	0.52
	EM	0.119±0.995	0.213±0.913	-0.70	198	0.49
	Regresyon	0.154±1.054	0.173±0.986	-0.13	198	0.90

Çizelge 4.39'da 200 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X2 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir..

Çizelge 4.39- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.125±0.922</b>	<b>-0.067±0.971</b>	<b>1.43</b>	<b>198</b>	<b>0.16</b>
%5 eksik	Pairwise	0.079±0.915	-0.053±0.981	0.95	188	0.34
	Ortalama	0.075±0.885	-0.050±0.962	0.95	198	0.34
	EM	0.071±0.887	-0.053±0.962	0.94	198	0.35
	Regresyon	0.046±0.906	-0.041±0.971	0.65	198	0.51
%10 eksik	Pairwise	0.137±0.931	-0.002±0.992	0.97	179	0.33
	Ortalama	0.135±0.916	0.008±0.914	0.98	198	0.33
	EM	0.136±0.916	0.003±0.916	1.02	198	0.31
	Regresyon	0.138±0.920	0.023±0.998	0.84	198	0.40
%15 eksik	Pairwise	0.080±0.908	-0.021±0.967	0.70	170	0.49
	Ortalama	0.070±0.927	-0.015±0.910	0.69	198	0.49
	EM	0.078±0.933	-0.011±0.915	0.71	198	0.48
	Regresyon	0.095±0.917	0.007±0.960	0.66	198	0.51
%20 eksik	Pairwise	0.112±0.928	-0.003±0.988	0.76	161	0.45
	Ortalama	0.097±0.817	0.004±0.909	0.76	198	0.45
	EM	0.089±0.826	0.018±0.919	0.58	198	0.57
	Regresyon	0.180±0.920	0.015±0.953	1.24	198	0.22

Çizelge 4.40'ta 200 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X3 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.40- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.025±0.960</b>	<b>0.106±1.084</b>	<b>-0.56</b>	<b>198</b>	<b>0.58</b>
%5 eksik	Pairwise	0.038±0.935	0.108±1.097	-0.47	188	0.64
	Ortalama	0.041±0.899	0.107±1.081	-0.47	198	0.64
	EM	0.061±0.903	0.104±1.082	-0.30	198	0.76
	Regresyon	0.017±0.958	0.084±1.092	-0.46	198	0.65
%10 eksik	Pairwise	-0.008±0.938	0.090±1.091	-0.65	178	0.52
	Ortalama	-0.002±0.881	0.086±1.044	-0.65	198	0.52
	EM	-0.024±0.885	0.081±1.045	-0.76	198	0.45
	Regresyon	-0.042±0.962	0.088±1.087	-0.89	198	0.37
%15 eksik	Pairwise	0.072±0.976	0.074±1.098	-0.01	171	0.99
	Ortalama	0.072±0.905	0.074±1.022	-0.01	198	0.99
	EM	0.074±0.909	0.073±1.026	0.01	198	0.99
	Regresyon	0.085±0.977	0.101±1.074	-0.11	198	0.91
%20 eksik	Pairwise	-0.071±0.956	0.077±1.103	-0.91	160	0.37
	Ortalama	-0.057±0.865	0.063±0.987	-0.91	198	0.37
	EM	-0.058±0.874	0.062±0.998	-0.90	198	0.37
	Regresyon	-0.085±0.937	0.121±1.124	-1.40	198	0.16

Çizelge 4.41'de 200 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X4 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.41- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.033±0.979</b>	<b>-0.009±1.019</b>	<b>-0.17</b>	<b>198</b>	<b>0.87</b>
%5 eksik	Pairwise	-0.037±1.000	0.012±1.024	-0.33	188	0.74
	Ortalama	-0.035±0.962	0.011±1.009	-0.33	198	0.74
	EM	-0.038±0.962	0.012±1.009	-0.36	198	0.72
	Regresyon	0.004±1.006	0.016±1.014	-0.08	198	0.94
%10 eksik	Pairwise	-0.046±0.980	-0.019±1.054	-0.18	180	0.86
	Ortalama	-0.045±0.937	-0.021±1.002	-0.18	198	0.86
	EM	-0.045±0.938	-0.018±1.004	-0.20	198	0.84
	Regresyon	-0.040±0.993	-0.047±1.025	0.05	198	0.96
%15 eksik	Pairwise	-0.041±1.029	-0.007±1.040	-0.30	168	0.76
	Ortalama	-0.036±0.924	0.005±0.979	-0.30	198	0.76
	EM	-0.038±0.926	0.001±0.981	-0.29	198	0.77
	Regresyon	-0.034±1.023	0.045±1.037	-0.54	198	0.59
%20 eksik	Pairwise	-0.061±1.015	0.003±1.020	-0.40	164	0.69
	Ortalama	-0.056±0.942	-0.035±0.912	-0.40	198	0.69
	EM	-0.059±0.944	0.006±0.916	-0.49	198	0.63
	Regresyon	-0.050±0.971	0.032±1.040	-0.58	198	0.57



Düşük korelasyonlu değişkenlere sahip 200 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde, farklı gruplara ait ortalamalar arasındaki farkların saptanmasında ise genel itibariyle Pairwise ve Ortalama atama yöntemleri daha benzer sonuçlar vermiştir. %10'luk eksik veri setlerinde, genel itibariyle Pairwise, Ortalama ve EM atama yöntemleri daha benzer sonuçlar vermiştir. %15'lik eksik veri setlerinde EM, genel itibariyle EM atama yöntemi daha benzer sonuçlar vermesine rağmen, genel olarak tüm yöntemlerden elde edilen anlamlılık düzeyleri (p) artmış ve t değerleri düşmüştür. %20'lik eksik veri setlerinde, ise genel itibariyle Pairwise ve Ortalama atama yöntemleri daha benzer sonuçlar verse de bir belirsizlik sözkonusudur.

#### 4.3.2- Yüksek Korelasyonlu 200 Birimlik Veri Seti

200 birim içeren tam veri setine ilişkin değişkenler arasındaki korelasyonları gösteren Çizelge aşağıdadır. Çizelge 4.42'de görüleceği gibi değişkenler arasındaki korelasyon katsayıları oldukça yüksektir ( $r > 0.85$ ).

Çizelge 4.42- 100 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri

	X1	X2	X3	X4
X1	1	0.89 (P<0.001)	0.89 (P<0.001)	0.89 (P<0.001)
X2		1	0.89 (P<0.001)	0.87 (P<0.001)
X3			1	0.89 (P<0.001)
X4				1

Not: Çizelgede parantez içindeki değerler olasılık değerleridir.

Çizelge 4.43'te 200 birimden oluşan düşük korelasyonlara sahip tam veri setinde sırasıyla %5, %10, %15 ve %20 değer eksiltilerek elde edilen veri setlerindeki eksik olma sayı ve yüzdeleri verilmiştir. Çizelge incelendiğinde eksiltmelerde verilen yüzde değerlerine yaklaşık azaltmalar gerçekleşmiştir. Bu farklılaşmalar tam rasgelelik koşullarının sağlanması için yerine koyma yöntemi ile eksiltme yönteminin kullanılmasından kaynaklanmaktadır. Ancak öngörülen yüzde eksiltme değerlerine yaklaşmıştır.

Çizelge 4.43- 200 birim içeren yüksek korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri

Değişken	Tam veri (n)	%5 Eksik		%10 Eksik		%15 Eksik		%20 Eksik	
		sayı	%	sayı	%	sayı	%	sayı	%
X1	200	10	5	19	9.5	28	14	38	19
X2	200	10	5	18	9	27	13.5	37	18.5
X3	200	10	5	18	9	27	13.5	38	19
X4	200	10	5	20	10	26	13	39	19.5
X5	200	0	0	0	0	0	0	0	0

Çizelge 4.44'de tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen ortalama ve standart sapma değerleri verilmiştir.

Çizelge 4.44- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri

		X1 $\bar{X} \pm s$	X2 $\bar{X} \pm s$	X3 $\bar{X} \pm s$	X4 $\bar{X} \pm s$
<b>Tam Veri</b>		<b>0.058±0.960</b>	<b>0.104±0.953</b>	<b>0.060±0.950</b>	<b>0.061±0.956</b>
%5 eksik	Listwise	0.060±0.944	0.100±0.948	0.051±0.920	0.057±0.954
	Pairwise	0.058±0.963	0.092±0.949	0.059±0.967	0.080±0.942
	Ortalama	0.058±0.939	0.092±0.925	0.059±0.942	0.080±0.918
	EM	0.062±0.961	0.100±0.962	0.065±0.951	0.070±0.958
	Regresyon	0.063±0.976	0.106±0.962	0.066±0.954	0.063±0.965
%10 eksik	Listwise	-0.003±0.988	0.093±0.978	-0.006±1.027	0.022±0.968
	Pairwise	0.060±0.983	0.100±0.928	0.019±0.963	0.088±0.973
	Ortalama	0.060±0.935	0.100±0.885	0.019±0.919	0.088±0.923
	EM	0.049±0.963	0.104±0.931	0.066±0.965	0.078±0.953
	Regresyon	-0.003±0.988	0.093±0.978	-0.006±1.027	0.022±0.968
%15 eksik	Listwise	0.062±0.984	0.158±0.973	0.064±0.941	0.093±0.954
	Pairwise	0.062±0.974	0.117±0.963	0.040±0.918	0.087±0.966
	Ortalama	0.062±0.903	0.117±0.985	0.040±0.83	0.087±0.900
	EM	0.074±0.979	0.115±0.957	0.057±0.942	0.066±0.953
	Regresyon	0.084±0.976	0.131±0.951	0.071±0.943	0.081±0.946
%20 eksik	Listwise	0.169±1.024	0.157±0.964	0.112±0.919	0.107±1.025
	Pairwise	0.119±0.960	0.060±0.966	0.109±0.931	0.033±0.981
	Ortalama	0.119±0.863	0.060±0.872	0.109±0.837	0.033±0.879
	EM	0.086±0.961	0.085±0.954	0.079±0.915	0.060±0.953
	Regresyon	0.093±0.940	0.083±0.944	0.075±0.918	0.042±0.965

Yüksek korelasyonlu değişkenlere sahip 200 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde, Regresyon atama yöntemi diğer atama yöntemlerine göre tam veri setinden elde edilen belirtici istatistiklere benzer sonuçlar

vermiştir ( $P>0.05$ ). %10'luk eksik veri setlerinde, EM atama yöntemi tam veri setinden elde edilen belirtici istatistiklere benzer sonuçlar vermiştir ( $P>0.05$ ). %15'lik eksik veri setlerinde Regresyon ve EM atama yöntemleri tam veri setinden elde edilen belirtici istatistiklere yakın sonuçlar vermiştir ( $P>0.05$ ). Türetilen 3 değişkende de bu sonuçlar benzerlik göstermektedir. %20'lik eksik veri setlerinde, EM atama yöntemi ile tam veri setinden elde edilen belirtici istatistikler benzerlik göstermektedir ( $P>0.05$ ).

Çizelge 4.45'te tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkeni bağımlı ve X2, X3 ve X4 değişkenleri bağımsız değişkenler olarak alındığında, elde edilen regresyon katsayıları, bu katsayıların anlamlılık düzeyleri ve belirleyicilik katsayıları ( $R^2$ ) verilmiştir.

Çizelge 4.45- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları

		Sabit ( $b_0$ )	$b_1$	$b_2$	$b_3$	$R^2$
	<b>Tam Veri</b>	<b>-0.015</b>	<b>0.340***</b>	<b>0.297***</b>	<b>0.335***</b>	<b>0.854</b>
%5 eksik	Listwise	-0.007	0.322***	0.319***	0.323***	0.848
	Ortalama	-0.010	0.325***	0.393***	0.238***	0.796
	EM	-0.014	0.342***	0.262***	0.362***	0.854
	Regresyon	-0.104***	-0.125	-0.608***	-0.612***	0.881
%10 eksik	Listwise	-0.044	0.402***	0.321***	0.247**	0.879
	Ortalama	-0.010	0.362***	0.281***	0.338***	0.788
	EM	-0.036	0.397***	0.267***	0.328***	0.880
	Regresyon	-0.033	0.394***	0.241***	0.350***	0.870
%15 eksik	Listwise	-0.043	0.359***	0.350**	0.279**	0.859
	Ortalama	-0.020	0.353***	0.257***	0.327***	0.711
	EM	-0.008	0.394***	0.269***	0.321***	0.852
	Regresyon	0.028	0.430***	0.162**	0.424***	0.869
%20 eksik	Listwise	0.036	0.419***	0.259*	0.350***	0.891
	Ortalama	0.059	0.278***	0.295***	0.345***	0.688
	EM	0.013	0.370***	0.285***	0.331***	0.864
	Regresyon	0.056*	0.197***	0.095	0.640***	0.864

\*  $p<0.05$       \*\*  $p<0.01$       \*\*\*  $p<0.001$

Yüksek korelasyonlu değişkenlere sahip 200 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setlerinde yapılan regresyon analizlerinde ise EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2=0.854$ ;  $R_T^2=0.854$ ). Regresyon atama yöntemiyle elde

edilen veri setinin belirleyicilik katsayısı olması gerekenden yüksek bulunmuştur ( $R_R^2=0.881$ ). %10'luk eksik veri setlerinde yapılan regresyon analizlerinde ise Listwise, EM ve Regresyon atama yöntemlerinden elde edilen belirleyicilik katsayıları benzer çıkmıştır ( $R_{L,EM,R}^2=0.88$ ;  $R_T^2=0.854$ ). %15'lik eksik veri yapılan regresyon analizlerinde ise EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2=0.852$ ;  $R_T^2=0.854$ ). Regresyon atama yöntemiyle elde edilen veri setinin belirleyicilik katsayısı olması gerekenden daha yüksek bulunmuştur ( $R_R^2=0.869$ ). %20'lik eksik veri setlerinde yapılan regresyon analizlerinde ise EM ve Regresyon atama yöntemlerinden elde edilen belirleyicilik katsayıları, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2=0.864$ ;  $R_R^2=0.864$ ;  $R_T^2=0.854$ ).

Çizelge 4.46'da 200 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.46- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.094±0.995</b>	<b>0.018±0.922</b>	<b>0.56</b>	<b>198</b>	<b>0.58</b>
%5 eksik	Pairwise	0.094±1.001	0.019±0.925	0.53	188	0.59
	Ortalama	0.091±0.963	0.020±0.915	0.54	198	0.59
	EM	0.103±0.990	0.016±0.916	0.64	198	0.52
	Regresyon	0.023±1.240	-0.022±1.063	0.27	198	0.79
%10 eksik	Pairwise	0.092±1.006	0.023±0.959	0.47	179	0.64
	Ortalama	0.089±0.958	0.026±0.911	0.47	198	0.64
	EM	0.087±0.987	0.005±0.923	0.60	198	0.55
	Regresyon	0.088±0.983	0.007±0.927	0.60	198	0.55
%15 eksik	Pairwise	0.100±0.980	0.017±0.971	0.56	170	0.58
	Ortalama	0.096±0.918	0.024±0.889	0.56	198	0.58
	EM	0.113±0.991	0.029±0.924	0.61	198	0.54
	Regresyon	-0.100±1.142	0.155±0.908	-1.73	198	0.09
%20 eksik	Pairwise	0.159±0.945	0.067±0.983	0.60	160	0.55
	Ortalama	0.153±0.875	0.080±0.852	0.60	198	0.55
	EM	0.125±0.968	0.041±0.900	0.64	198	0.53
	Regresyon	0.157±0.966	0.166±0.822	-0.07	198	0.95

Çizelge 4.47’de 200 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X2 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.47- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.152±1.012</b>	<b>0.050±0.882</b>	<b>0.75</b>	<b>198</b>	<b>0.45</b>
%5 eksik	Pairwise	0.105±1.009	0.077±0.881	0.20	188	0.84
	Ortalama	0.105±0.980	0.078±0.862	0.20	198	0.84
	EM	0.147±1.019	0.046±0.880	0.74	198	0.46
	Regresyon	0.169±1.019	0.070±0.692	0.79	198	0.43
%10 eksik	Pairwise	0.147±1.010	0.045±0.822	0.74	180	0.46
	Ortalama	0.143±0.971	0.051±0.776	0.74	198	0.46
	EM	0.159±0.994	0.041±0.838	0.90	198	0.37
	Regresyon	0.157±1.004	0.060±0.859	0.73	198	0.47
%15 eksik	Pairwise	0.210±1.027	0.016±0.883	1.32	171	0.19
	Ortalama	0.195±0.942	0.027±0.834	1.32	198	0.19
	EM	0.169±0.993	0.054±0.872	0.86	198	0.39
	Regresyon	-0.112±1.056	0.145±0.829	-1.89	198	0.06
%20 eksik	Pairwise	0.130±1.019	-0.017±0.906	0.97	161	0.34
	Ortalama	0.115±0.907	-0.043±0.829	0.97	198	0.33
	EM	0.136±0.986	0.027±0.857	0.83	198	0.41
	Regresyon	0.129±0.994	0.183±0.895	-0.40	198	0.69

Çizelge 4.48’de 200 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X3 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.48- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.075±1.049</b>	<b>0.043±0.826</b>	<b>0.24</b>	<b>198</b>	<b>0.81</b>
%5 eksik	Pairwise	0.096±1.067	0.016±0.840	0.56	188	0.57
	Ortalama	0.094±1.042	0.019±0.817	0.56	198	0.57
	EM	0.090±1.044	0.035±0.822	0.41	198	0.68
	Regresyon	0.115±1.122	0.044±0.857	0.50	198	0.62
%10 eksik	Pairwise	0.025±1.071	0.013±0.833	0.08	180	0.93
	Ortalama	0.024±1.014	0.013±0.801	0.08	198	0.93
	EM	0.081±1.056	0.049±0.836	0.23	198	0.82
	Regresyon	0.060±1.045	0.045±0.827	0.11	198	0.91
%15 eksik	Pairwise	0.080±0.995	-0.007±0.821	0.62	171	0.54
	Ortalama	0.075±0.932	-0.0001±0.756	0.62	198	0.54
	EM	0.098±1.014	0.010±0.813	0.66	198	0.51
	Regresyon	-0.075±1.075	0.118±0.962	-1.33	198	0.19
%20 eksik	Pairwise	0.163±1.013	0.049±0.832	0.77	160	0.44
	Ortalama	0.151±0.902	0.059±0.757	0.78	198	0.44
	EM	0.102±0.972	0.051±0.972	0.41	198	0.69
	Regresyon	0.132±1.043	0.187±0.811	-0.41	198	0.68

Çizelge 4.49'da 200 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X4 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.49- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.152±0.977</b>	<b>-0.044±0.924</b>	<b>1.45</b>	<b>198</b>	<b>0.15</b>
%5 eksik	Pairwise	0.165±0.951	-0.019±0.928	1.35	188	0.18
	Ortalama	0.161±0.928	-0.014±0.903	1.34	198	0.18
	EM	0.145±0.982	-0.018±0.916	1.21	198	0.23
	Regresyon	0.146±1.068	0.085±0.893	0.43	198	0.67
%10 eksik	Pairwise	0.153±0.994	0.001±0.944	1.04	178	0.30
	Ortalama	0.151±0.975	0.016±0.859	1.03	198	0.31
	EM	0.159±0.978	-0.017±0.901	1.32	198	0.19
	Regresyon	0.156±0.976	-0.017±0.919	1.29	198	0.20
%15 eksik	Pairwise	0.219±0.994	-0.072±0.912	1.99	172	0.048*
	Ortalama	0.204±0.937	-0.048±0.841	1.99	198	0.048*
	EM	0.167±0.972	-0.050±0.885	1.65	198	0.10
	Regresyon	-0.164±1.120	0.111±0.939	-1.86	198	0.06
%20 eksik	Pairwise	0.092±1.007	-0.039±0.951	0.84	159	0.40
	Ortalama	0.082±0.912	-0.024±0.842	0.84	198	0.40
	EM	0.141±0.960	-0.034±0.884	1.34	198	0.18
	Regresyon	0.109±0.997	0.075±0.834	0.27	198	0.79

Yüksek korelasyonlu değişkenlere sahip 200 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise genel itibariyle EM ve Ortalama ve Pairwise atama yöntemleri tam veri seti ile benzer sonuçlar vermektedir. %10'luk eksik veri setlerinde genel itibariyle EM atama yöntemi benzer sonuçlar vermiştir. %15'lik eksik veri setlerinde iki gruba ait ortalamalar arasındaki farkların saptanmasında ise genel itibariyle EM atama yöntemi tam veri setinden elde edilen değerlere benzer sonuçlar vermiştir. X4 değişkeninde, tam veri setinde ortalamalar arasında fark bulunmazken ( $t=1.45$ ;  $p=0.15$ ), Pairwise ve Ortalama atama yöntemleri sonucunda elde edilen veri setinde ortalamalar arasındaki fark anlamlı bulunmuştur ( $t_p=1.99$ ;  $p=0,048^*$ ;  $t_o=1.99$ ;  $p=0.048^*$ ). %20'lik eksik veri setlerinde ise genel itibariyle EM atama yöntemi tam veri setinden elde edilen değerlere benzer sonuçlar vermiştir.

#### 4.4- 300 Birimlik Türetilmiş Veri Setleri

Düşük ve yüksek korelasyonlara sahip 300 birimlik veri setlerine ilişkin analizler aşağıdaki bölümlerde verilmiştir.

##### 4.4.1- Düşük Korelasyonlu 300 Birimlik Veri Seti

300 birim içeren tam veri setine ilişkin değişkenler arasındaki korelasyonları gösteren çizelge 4.50'de görüleceği gibi değişkenler arasındaki korelasyon katsayıları oldukça düşüktür ( $r < 0,20$ ).

Çizelge 4.50- 300 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri

	X1	X2	X3	X4
X1	1	0,13 (P<0,05)	0,09 (P>0,05)	0,02 (P>0,05)
X2		1	-0,01 (P>0,05)	0,11 (P>0,05)
X3			1	0,06 (P>0,05)
X4				1

Not: Çizelgede parantez içindeki değerler olasılık değerleridir.

Çizelge 4.51'de 300 birimden oluşan düşük korelasyonlara sahip tam veri setinde sırasıyla %5, %10, %15 ve %20 değer eksiltilerek elde edilen veri setlerindeki eksik olma sayı ve yüzdeleri verilmiştir. Çizelge incelendiğinde eksiltmelerde verilen yüzde değerlerine yaklaşık azaltmalar gerçekleşmiştir. Bu farklılaşmalar tam rasgelelik koşullarının sağlanması için yerine koyma yöntemi ile eksiltme yönteminin kullanılmasından kaynaklanmaktadır. Ancak öngörülen yüzde eksiltme değerlerine yaklaşılmıştır.

Çizelge 4.51- 300 birim içeren düşük korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri

Değişken	Tam veri (n)	%5 Eksik		%10 Eksik		%15 Eksik		%20 Eksik	
		sayı	%	sayı	%	sayı	%	sayı	%
X1	300	15	5	29	9.7	41	13.7	54	18
X2	300	15	5	29	9.7	41	13.7	55	18.3
X3	300	15	5	28	9.3	42	14	57	19
X4	300	15	5	28	9.3	44	14.7	56	18.7
X5	300	0	0	0	0	0	0	0	0

Çizelge 4.52'de tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen ortalama ve standart sapma değerleri verilmiştir.

Çizelge 4.52- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri

		<b>X1</b> $\bar{X} \pm S$	<b>X2</b> $\bar{X} \pm S$	<b>X3</b> $\bar{X} \pm S$	<b>X4</b> $\bar{X} \pm S$
	<b>Tam Veri</b>	<b>0.021±0.935</b>	<b>0.014±1.037</b>	<b>-0.015±1.064</b>	<b>-0.057±0.976</b>
%5 eksik	Listwise	-0.008±0.930	-0.031±1.019	-0.013±1.078	-0.066±0.978
	Pairwise	-0.001±0.939	0.007±1.021	-0.026±1.062	-0.053±0.977
	Ortalama	-0.001±0.915	0.007±0.995	-0.026±1.035	-0.053±0.952
	EM	-0.003±0.939	0.006±1.021	-0.025±1.062	-0.051±0.977
	Regresyon	0.004±0.937	0.011±1.022	-0.040±1.054	-0.046±0.975
%10 eksik	Listwise	0.027±0.940	0.096±1.070	-0.057±1.080	-0.023±0.941
	Pairwise	0.032±0.935	0.040±1.035	-0.012±1.068	-0.069±0.982
	Ortalama	0.032±0.889	0.040±0.984	-0.012±1.017	-0.069±0.935
	EM	0.033±0.935	0.041±1.036	-0.012±1.068	-0.069±0.982
	Regresyon	0.009±0.934	0.023±1.035	-0.004±1.053	-0.066±0.972
%15 eksik	Listwise	0.106±0.964	0.005±1.010	-0.035±1.087	-0.015±0.984
	Pairwise	0.032±0.949	0.071±1.003	-0.072±1.064	-0.035±0.971
	Ortalama	0.032±0.882	0.071±0.932	-0.072±0.987	-0.035±0.897
	EM	0.033±0.949	0.071±1.003	-0.071±1.064	-0.035±0.971
	Regresyon	0.073±0.957	0.070±1.000	-0.053±1.071	-0.038±0.956
%20 eksik	Listwise	-0.030±0.944	-0.065±1.080	-0.051±1.081	-0.031±1.033
	Pairwise	0.034±0.952	-0.019±1.046	-0.052±1.07	-0.056±0.981
	Ortalama	0.034±0.861	-0.019±0.945	-0.052±0.996	-0.056±0.884
	EM	-0.032±0.951	-0.020±1.046	-0.052±1.107	-0.055±0.981
	Regresyon	0.021±0.935	-0.008±1.019	-0.026±1.024	-0.062±1.003

Düşük korelasyonlu değişkenlere sahip 300 birimlik veri setlerinin analizleri sonucunda; %5'lik eksik veri setinde, EM atama yöntemleri tam veri setinden elde edilen belirtici istatistiklere benzer sonuçlar vermiştir ( $P>0.05$ ). %10'luk eksik veri setlerinde aynı dağılım yapısına sahip 4 farklı değişken olmasına rağmen farklı sonuçlar elde edilmiştir. Genel olarak benzer ortalamalar Pairwise, Ortalama ve EM atama sonucunda elde edilen veri setinden elde edilmiştir. %15'lik eksik veri setinin tamamlanmasında, Pairwise, Ortalama ve EM atama yöntemi hemen hemen aynı sonuçları vermiştir ( $P>0.05$ ). Fakat farklı atama yöntemlerinden elde edilen ortalama ve standart sapma değerleri, tam veri setinden elde edilen değerlerden farklı bulunmuştur ( $P<0.05$ ). %20'lik eksik veri setinin tamamlanmasında ise Regresyon atama yöntemi ile tam veri setinden elde edilen belirtici istatistikler arasında benzerlik gözlenmiştir ( $P>0.05$ ).



Çizelge 4.53'te tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden oluşan yeni veri setlerinden elde edilmiş regresyon katsayıları, bu katsayıların anlamlılık düzeyleri ve belirleyicilik katsayıları ( $R^2$ ) verilmiştir.

Çizelge 4.53- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları

		Sabit ( $b_0$ )	$b_1$	$b_2$	$b_3$	$R^2$
	<b>Tam Veri</b>	<b>0.020</b>	<b>0.114*</b>	<b>0.077</b>	<b>-0.001</b>	<b>0.023</b>
%5 eksik	Listwise	-0.004	0.060	0.054	0.016	0.009
	Ortalama	0.0002	0.106*	0.074	0.010	0.021
	EM	-0.001	0.105*	0.075	0.011	0.022
	Regresyon	-0.137**	0.122*	0.014	0.212**	0.056
%10 eksik	Listwise	0.017	0.126*	0.039	-0.001	0.023
	Ortalama	0.030	0.086	0.063	0.003	0.015
	EM	0.030	0.087	0.063	0.005	0.015
	Regresyon	-0.095	-0.233***	-0.093	-0.043	0.048
%15 eksik	Listwise	0.100	0.083	0.098	0.019	0.022
	Ortalama	0.031	0.087	0.077	0.065	0.015
	EM	0.033	0.087	0.081	0.007	0.016
	Regresyon	0.086	0.080	0.014	0.083	0.013
%20 eksik	Listwise	-0.018	0.0142	0.054	0.017	0.031
	Ortalama	0.038	0.117*	0.074	-0.040	0.024
	EM	0.036	0.121*	0.080	-0.042	0.026
	Regresyon	0.023	0.149**	-0.012	-0.161**	0.048

\*  $p < 0.05$       \*\*  $p < 0.01$       \*\*\*  $p < 0.001$

Düşük korelasyonlu değişkenlere sahip 300 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde yapılan regresyon analizlerinde EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2 = 0.022$ ;  $R_T^2 = 0.023$ ). %10'luk eksik veri setlerinde yapılan analizlerde Listwise, Ortalama ve EM atama yöntemlerinden elde edilen belirleyicilik katsayıları, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{L,O,EM}^2 = 0.02$ ;  $R_T^2 = 0.023$ ). %15'lik eksik veri setlerinde yapılan regresyon analizlerinde Listwise atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_L^2 = 0.022$ ;  $R_T^2 = 0.023$ ). %20'lik eksik veri setlerinde yapılan regresyon analizlerinde ise EM ve Ortalama atama yöntemlerinden elde edilen belirleyicilik katsayıları, tam veri setinden elde edilen değere daha yakın bulunmuştur ( $R_L^2 = 0.026$ ;  $R_O^2 = 0.024$   $R_T^2 = 0.023$ ).

Çizelge 4.54'te 300 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.54- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.017±0.922</b>	<b>0.024±0.950</b>	<b>-0.06</b>	<b>298</b>	<b>0.95</b>
%5 eksik	Pairwise	-0.018±0.913	0.014±0.966	-0.29	283	0.77
	Ortalama	-0.017±0.887	0.014±0.943	-0.29	298	0.77
	EM	-0.018±0.887	0.012±0.943	-0.29	298	0.78
	Regresyon	-0.094±0.880	-0.116±0.888	0.22	298	0.83
%10 eksik	Pairwise	0.038±0.914	0.026±0.957	0.10	269	0.92
	Ortalama	0.038±0.863	0.027±0.915	0.10	298	0.92
	EM	0.037±0.863	0.029±0.916	0.08	298	0.94
	Regresyon	0.042±0.958	-0.149±0.946	1.74	298	0.08
%15 eksik	Pairwise	0.009±0.931	0.054±0.969	-0.38	257	0.71
	Ortalama	0.012±0.868	0.050±0.897	-0.38	298	0.71
	EM	0.013±0.869	0.053±0.897	-0.39	298	0.70
	Regresyon	0.047±0.958	0.166±0.916	-1.11	298	0.27
%20 eksik	Pairwise	0.017±0.938	0.048±0.966	-0.26	244	0.80
	Ortalama	0.021±0.821	0.046±0.901	-0.26	298	0.80
	EM	0.017±0.821	0.046±0.901	-0.29	298	0.77
	Regresyon	0.030±0.889	0.074±0.950	-0.42	298	0.68

Çizelge 4.55'te 300 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X2 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.55- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.029±1.001</b>	<b>0.055±1.071</b>	<b>-0.70</b>	<b>298</b>	<b>0.49</b>
%5 eksik	Pairwise	-0.035±0.999	0.046±1.043	-0.67	283	0.51
	Ortalama	-0.032±0.971	0.044±1.019	-0.67	298	0.51
	EM	-0.034±0.971	0.044±1.019	-0.68	298	0.50
	Regresyon	-0.032±0.951	-0.076±0.904	0.40	298	0.69
%10 eksik	Pairwise	-0.044±1.002	0.122±1.065	-1.32	269	0.19
	Ortalama	-0.037±0.960	0.113±1.004	-1.32	298	0.19
	EM	-0.035±0.960	0.114±1.004	-1.31	298	0.19
	Regresyon	-0.170±0.828	-0.088±0.935	-0.80	298	0.43
%15 eksik	Pairwise	0.019±1.003	0.118±1.004	-0.79	257	0.43
	Ortalama	0.027±0.921	0.112±0.943	-0.79	298	0.43
	EM	0.069±0.921	0.112±0.944	-0.79	298	0.43
	Regresyon	0.228±0.981	0.261±0.898	-0.31	298	0.76
%20 eksik	Pairwise	0.001±1.014	-0.039±1.081	0.29	243	0.77
	Ortalama	-0.027±0.930	-0.035±0.961	0.29	298	0.77
	EM	-0.009±0.931	-0.030±0.963	0.19	298	0.85
	Regresyon	0.057±0.978	0.185±0.997	-1.12	298	0.27

Çizelge 4.56'da 300 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X3 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.56- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.015±0.977</b>	<b>-0.044±1.143</b>	<b>0.48</b>	<b>298</b>	<b>0.63</b>
%5 eksik	Pairwise	-0.001±0.982	-0.049±0.136	0.38	283	0.70
	Ortalama	-0.021±0.958	-0.048±1.106	0.38	298	0.70
	EM	-0.001±0.958	-0.048±1.106	0.39	298	0.70
	Regresyon	-0.223±0.921	-0.065±1.043	-1.38	298	0.17
%10 eksik	Pairwise	0.067±0.976	-0.087±1.148	1.19	270	0.24
	Ortalama	0.060±0.932	-0.080±1.090	1.19	298	0.24
	EM	0.060±0.933	-0.080±1.090	1.19	298	0.23
	Regresyon	0.139±0.942	0.254±1.048	-1.00	298	0.32
%15 eksik	Pairwise	-0.058±0.992	-0.084±1.129	0.20	256	0.84
	Ortalama	-0.060±0.906	-0.083±1.060	0.20	298	0.84
	EM	-0.055±0.907	0.087±1.061	0.28	298	0.78
	Regresyon	-0.042±0.962	-0.106±1.095	0.54	298	0.59
%20 eksik	Pairwise	-0.051±1.017	-0.053±1.189	0.01	241	0.99
	Ortalama	-0.051±0.910	-0.053±1.075	0.01	298	0.99
	EM	-0.054±0.911	-0.050±1.076	-0.03	298	0.97
	Regresyon	-0.115±1.138	-0.038±1.196	-0.57	298	0.57

Çizelge 4.57'de 300 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X4 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.57- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.138±1.059</b>	<b>0.021±0.887</b>	<b>-1.41</b>	<b>298</b>	<b>0.16</b>
%5 eksik	Pairwise	-0.136±1.077	0.028±0.866	-1.42	283	0.16
	Ortalama	-0.133±1.055	0.023±0.840	-1.42	298	0.16
	EM	-0.133±1.055	0.026±0.840	-1.45	298	0.15
	Regresyon	0.116±0.915	0.257±0.825	-1.41	298	0.16
%10 eksik	Pairwise	-0.145±1.069	0.005±0.886	-1.26	270	0.21
	Ortalama	0.139±1.028	-0.030±0.836	-1.26	298	0.21
	EM	-0.136±1.028	-0.005±0.837	-1.21	298	0.21
	Regresyon	-0.383±1.004	-0.080±1.025	-2.58	298	0.01*
%15 eksik	Pairwise	-0.114±1.033	0.039±0.906	-1.27	254	0.21
	Ortalama	-0.102±0.952	0.029±0.839	-1.27	298	0.21
	EM	-0.100±0.953	0.027±0.840	-1.22	298	0.22
	Regresyon	0.051±0.863	0.038±0.849	0.13	298	0.90
%20 eksik	Pairwise	-0.124±1.070	0.006±0.891	-1.04	242	0.30
	Ortalama	-0.110±0.957	-0.046±0.809	-1.04	298	0.30
	EM	-0.112±0.958	-0.002±0.809	-1.07	298	0.28
	Regresyon	-0.055±1.095	-0.071±0.916	0.14	298	0.89

Düşük korelasyonlu değişkenlere sahip 300 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde, farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise genel itibariyle EM atama yöntemi daha benzer sonuçlar vermiştir. %10'luk eksik veri setlerinde genel itibariyle EM atama yöntemi daha benzer sonuçlar vermiştir. X4 değişkeninde, tam veri setinde ortalamalar arasında fark bulunmazken ( $t=0,41$ ;  $p=0.16$ ), Regresyon atama yöntemi sonucunda elde edilen veri setinde ortalamalar arasındaki fark anlamlı bulunmuştur ( $t=-2,58$ ;  $p<0.01^*$ ) (Çizelge 5.57). %15'lik eksik veri setlerinde, farklı iki gruba ait ortalamaların karşılaştırılmasında genel itibariyle Listwise, Ortalama ve EM atama yöntemleri benzer sonuçlar vermiştir. %20'lik eksik veri setlerinde, genel itibariyle Regresyon atama yöntemi daha benzer sonuçlar vermiştir.

#### 4.4.2- Yüksek Korelasyonlu 300 Birimlik Veri Seti

300 birim içeren veri setine ilişkin değişkenler arasındaki korelasyonları gösteren Çizelge aşağıdadır. Çizelge 4.58'de görüleceği gibi değişkenler arasındaki korelasyon katsayıları oldukça yüksektir ( $r>0.85$ ).

Çizelge 4.58- 300 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri

	X1	X2	X3	X4
X1	1	0.90 (P<0.001)	0.89 (P<0.001)	0.90 (P<0.001)
X2		1	0.90 (P<0.001)	0.90 (P<0.001)
X3			1	0.89 (P<0.001)
X4				1

Not: Çizelgede parantez içindeki değerler olasılık değerleridir.

Çizelge 4.59'da 300 birimden oluşan yüksek korelasyonlara sahip tam veri setinde sırasıyla %5, %10, %15 ve %20 değer eksiltilerek elde edilen veri setlerindeki eksik olma sayı ve yüzdeleri verilmiştir. Çizelge incelendiğinde eksiltmelerde verilen yüzde değerlerine yaklaşık azaltmalar gerçekleşmiştir. Bu farklılaşmalar tam rasgelelik koşullarının sağlanması için yerine koyma yöntemi ile eksiltme yönteminin kullanılmasından kaynaklanmaktadır. Ancak öngörülen yüzde eksiltme değerlerine yaklaşılmıştır.

Çizelge 4.59- 300 birim içeren yüksek korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayısı ve yüzdeleri

Değişken	Tam veri (n)	%5 Eksik		%10 Eksik		%15 Eksik		%20 Eksik	
		sayı	%	sayı	%	sayı	%	sayı	%
X1	300	15	5	30	10	39	13	53	17.7
X2	300	15	5	27	9	44	14.7	55	18.3
X3	300	15	5	29	9.7	45	15	55	18.3
X4	300	15	5	28	9.3	42	14	55	18.3
X5	300	0	0	0	0	0	0	0	0

Çizelge 4.60'ta tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen ortalama ve standart sapma değerleri verilmiştir.

Çizelge 4.60- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri

		X1 $\bar{X} \pm S$	X2 $\bar{X} \pm S$	X3 $\bar{X} \pm S$	X4 $\bar{X} \pm S$
<b>Tam Veri</b>		<b>0.006±1.006</b>	<b>-0.005±0.979</b>	<b>-0.007±0.998</b>	<b>-0.010±0.982</b>
%5 eksik	Listwise	-0.015±1.029	-0.045±1.012	-0.036±1.022	-0.043±1.016
	Pairwise	0.019±1.011	-0.005±0.989	-0.016±1.001	-0.027±0.984
	Ortalama	0.019±0.986	-0.005±0.964	-0.016±0.976	-0.027±0.959
	EM	0.016±1.004	-0.009±0.993	-0.013±0.994	-0.010±0.979
	Regresyon	0.014±1.009	-0.008±0.980	-0.014±0.990	-0.016±0.974
%10 eksik	Listwise	-0.005±1.007	-0.018±1.005	0.004±1.023	-0.001±0.980
	Pairwise	-0.005±1.001	-0.008±0.984	-0.017±1.009	0.005±0.978
	Ortalama	-0.005±0.949	-0.008±0.939	-0.017±0.958	0.005±0.931
	EM	-0.001±1.001	-0.025±0.985	-0.007±1.001	-0.011±0.976
	Regresyon	-0.009±0.999	-0.019±0.976	-0.007±1.004	-0.002±0.975
%15 eksik	Listwise	0.031±0.012	0.065±1.029	0.030±1.016	0.061±1.004
	Pairwise	0.012±1.019	0.041±0.985	-0.018±0.994	-0.008±1.010
	Ortalama	0.012±0.951	0.041±0.909	-0.018±0.917	-0.008±0.936
	EM	0.001±1.011	-0.015±0.995	-0.013±0.994	0.007±0.981
	Regresyon	-0.005±1.014	0.014±0.968	-0.006±0.991	-0.006±0.994
%20 eksik	Listwise	0.030±1.043	0.015±1.048	0.024±1.024	0.014±1.029
	Pairwise	0.009±1.017	-0.048±1.002	0.002±1.003	-0.032±0.982
	Ortalama	0.009±0.923	-0.048±0.906	0.002±0.906	-0.032±0.887
	EM	0.002±1.012	-0.016±0.999	-0.007±0.991	-0.015±0.981
	Regresyon	0.011±1.008	-0.014±0.983	-0.005±0.994	-0.002±0.971

Yüksek korelasyonlu değişkenlere sahip 300 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde, EM atama yöntemi tam veri setinden elde edilen

belirtici istatistiklere benzer sonuçlar vermiştir ( $P>0.05$ ). %10'luk eksik veri setlerinde Pairwise, Ortalama ve EM atama yöntemleri tam veri setinden elde edilen belirtici istatistiklere yakın sonuçlar vermiştir ( $P>0.05$ ). %15'lik eksik veri setlerinde, Regresyon atama yöntemi tam veri setinden elde edilen belirtici istatistiklere yakın sonuçlar vermiştir ( $P>0.05$ ). %20'lik eksik veri setlerinde, EM atama yöntemi ile tam veri setinden elde edilen belirtici istatistikler benzerlik göstermektedir ( $P>0.05$ ).

Çizelge 4.61'de tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen yeni veri setlerinden elde edilmiş regresyon katsayıları, bu katsayıların anlamlılık düzeyleri ve belirleyicilik katsayıları ( $R^2$ ) verilmiştir.

Çizelge 4.61- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları

		Sabit ( $b_0$ )	$b_1$	$b_2$	$b_3$	$R^2$
	<b>Tam Veri</b>	<b>0.013</b>	<b>0.316***</b>	<b>0.274***</b>	<b>0.392***</b>	<b>0.863</b>
%5 eksik	Listwise	0.025	0.346***	0.281***	0.347***	0.867
	Ortalama	0.03	0.450***	0.285***	0.243***	0.839
	EM	0.027	0.371***	0.259***	0.346***	0.862
	Regresyon	-0.008	0.670***	0.433***	-0.027	0.893
%10 eksik	Listwise	-0.001	0.317***	0.284***	0.366***	0.865
	Ortalama	0.016	0.314***	0.322***	0.293***	0.752
	EM	0.014	0.301***	0.301***	0.369***	0.861
	Regresyon	0.055*	0.181**	0.092	0.736***	0.820
%15 eksik	Listwise	-0.019	0.293***	0.299***	0.367***	0.866
	Ortalama	0.015	0.153**	0.316***	0.468***	0.724
	EM	0.006	0.288***	0.301***	0.389***	0.860
	Regresyon	-0.015	0.252***	0.519***	0.238**	0.810
%20 eksik	Listwise	0.013	0.321***	0.296***	0.363***	0.882
	Ortalama	0.0037	0.398***	0.255***	0.270***	0.682
	EM	0.015	0.294***	0.311***	0.367***	0.857
	Regresyon	0.023	0.532***	0.337***	0.106	0.827

\*  $p<0.05$       \*\*  $p<0.01$       \*\*\*  $p<0.001$

Yüksek korelasyonlu değişkenlere sahip 300 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde yapılan regresyon analizlerinde ise EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2=0.862$ ;  $R_T^2=0.863$ ). Regresyon atama yöntemiyle elde edilen veri setinin belirleyicilik katsayısı olması gerekenden yüksek bulunmuştur

( $R_R^2=0.893$ ). %10'lük eksik veri setlerinde EM ve Listwise atama yöntemlerinden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2=0.861$ ;  $R_R^2=0.863$ ;  $R_T^2=0.863$ ). %15'lik eksik veri setlerinde, EM ve Listwise atama yöntemlerinden elde edilen belirleyicilik katsayıları, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2=0.860$ ;  $R_L^2=0.866$ ;  $R_T^2=0.863$ ). %20'lik eksik veri setlerinde ise EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_{EM}^2=0.857$ ;  $R_T^2=0.863$ ).

Çizelge 4.62'de 300 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.62- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.067±1.005</b>	<b>0.071±1.006</b>	<b>-1.19</b>	<b>298</b>	<b>0.24</b>
%5 eksik	Pairwise	-0.067±1.006	0.101±1.014	-1.40	283	0.16
	Ortalama	-0.064±0.991	0.095±0.977	-1.40	298	0.16
	EM	-0.064±1.004	0.089±0.990	-1.33	298	0.19
	Regresyon	0.293±0.871	0.141±0.909	1.48	298	0.14
%10 eksik	Pairwise	-0.085±0.996	0.068±1.003	-1.26	268	0.21
	Ortalama	-0.078±0.946	0.061±0.950	-1.26	298	0.21
	EM	-0.068±0.989	0.061±0.988	-1.13	298	0.26
	Regresyon	0.106±1.039	0.085±0.974	0.18	298	0.86
%15 eksik	Pairwise	-0.099±1.008	0.111±1.023	-1.67	259	0.10
	Ortalama	-0.084±0.935	0.099±0.959	-1.67	298	0.10
	EM	-0.089±0.990	0.082±0.997	-1.49	298	0.14
	Regresyon	-0.047±0.986	-0.095±0.974	0.43	298	0.67
%20 eksik	Pairwise	-0.074±1.043	0.087±0.990	-1.25	245	0.21
	Ortalama	-0.060±0.951	0.073±0.894	-1.25	298	0.21
	EM	-0.059±1.005	0.058±0.973	-1.03	298	0.30
	Regresyon	0.079±0.974	0.073±1.051	0.05	298	0.96

Çizelge 4.63'te 300 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X2 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.63- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.061±0.978</b>	<b>0.046±0.981</b>	<b>-0.95</b>	<b>298</b>	<b>0.34</b>
%5 eksik	Pairwise	-0.070±0.986	0.055±0.992	-1.07	283	0.29
	Ortalama	-0.068±0.969	0.052±0.960	-1.07	298	0.28
	EM	-0.072±0.983	0.047±0.992	-1.04	298	0.30
	Regresyon	0.276±0.728	0.183±0.813	1.04	298	0.30
%10 eksik	Pairwise	-0.067±0.973	0.044±0.995	-0.94	271	0.35
	Ortalama	-0.062±0.924	0.040±0.953	-0.94	298	0.35
	EM	-0.087±0.974	0.032±0.975	-1.06	298	0.29
	Regresyon	0.041±0.914	0.027±0.850	0.14	298	0.89
%15 eksik	Pairwise	-0.009±0.998	0.086±0.974	-0.77	254	0.44
	Ortalama	-0.001±0.918	0.080±0.903	-0.77	298	0.44
	EM	-0.060±0.963	0.027±0.993	-0.77	298	0.44
	Regresyon	-0.003±0.863	-0.103±0.871	1.00	298	0.32
%20 eksik	Pairwise	-0.074±1.022	-0.025±0.989	-0.38	243	0.71
	Ortalama	-0.068±0.904	-0.029±0.910	-0.38	298	0.71
	EM	-0.072±0.972	0.035±0.983	-0.95	298	0.34
	Regresyon	0.067±0.902	0.028±1.003	0.35	298	0.73

Çizelge 4.64'te 300 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X3 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.64- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.056±0.957</b>	<b>0.037±1.036</b>	<b>-0.80</b>	<b>298</b>	<b>0.42</b>
%5 eksik	Pairwise	-0.048±0.962	0.014±1.040	-0.52	283	0.61
	Ortalama	-0.047±0.952	0.011±0.999	-0.52	298	0.61
	EM	-0.049±0.954	0.020±1.020	-0.61	298	0.55
	Regresyon	0.239±0.847	0.102±0.894	1.36	298	0.17
%10 eksik	Pairwise	-0.097±0.960	0.053±1.048	-1.23	269	0.22
	Ortalama	-0.088±0.905	0.047±1.003	-1.23	298	0.22
	EM	-0.050±0.959	0.032±1.021	-0.72	298	0.47
	Regresyon	0.046±0.908	0.042±0.898	0.04	298	0.97
%15 eksik	Pairwise	-0.073±0.957	0.032±1.028	-0.84	253	0.40
	Ortalama	-0.065±0.884	0.024±0.946	-0.84	298	0.40
	EM	-0.051±0.962	0.021±0.995	-0.63	298	0.53
	Regresyon	0.010±0.885	-0.091±0.930	0.96	298	0.34
%20 eksik	Pairwise	-0.013±0.972	0.015±1.033	-0.21	243	0.83
	Ortalama	-0.099±0.871	-0.012±0.940	-0.21	298	0.83
	EM	-0.057±0.930	0.039±1.005	-0.85	298	0.39
	Regresyon	0.070±0.947	0.059±1.109	0.09	298	0.93

Çizelge 4.65'de 300 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X4 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.



Çizelge 4.65- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.047±0.978</b>	<b>0.024±0.988</b>	<b>-0.63</b>	<b>298</b>	<b>0.53</b>
%5 eksik	Pairwise	-0.061±0.985	0.004±0.986	-0.56	283	0.58
	Ortalama	-0.059±0.961	0.003±0.960	-0.56	298	0.58
	EM	-0.041±0.977	0.018±0.974	-0.53	298	0.60
	Regresyon	0.235±0.771	0.041±0.818	2.11	298	0.04*
%10 eksik	Pairwise	-0.045±0.960	0.051±0.997	-0.81	270	0.42
	Ortalama	-0.041±0.922	0.046±0.941	-0.81	298	0.42
	EM	-0.053±0.959	0.027±0.975	-0.72	298	0.48
	Regresyon	0.052±0.998	0.030±0.896	0.21	298	0.84
%15 eksik	Pairwise	-0.073±1.013	0.050±1.007	-0.98	256	0.33
	Ortalama	-0.063±0.931	0.043±0.941	-0.98	298	0.33
	EM	-0.033±0.970	0.044±0.964	-0.69	298	0.49
	Regresyon	-0.055±0.970	-0.115±0.968	0.54	298	0.59
%20 eksik	Pairwise	-0.036±1.024	-0.028±0.947	-0.06	243	0.95
	Ortalama	-0.035±0.921	-0.029±0.858	-0.06	298	0.95
	EM	-0.051±0.950	0.017±0.970	-0.61	298	0.54
	Regresyon	0.107±0.868	0.029±0.978	0.72	298	0.47

Yüksek korelasyonlu değişkenlere sahip 300 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise genel itibariyle EM atama yöntemi tam veri seti ile benzer sonuçlar vermiştir. Regresyon atama yönteminde ortalamalar arasındaki farkın araştırılması aşamasında X3 değişkeni anlamlı derecede farklı bulunmuştur. tam veri setinde X3 değişkenine ait ortalamalar arasında fark bulunmazken ( $t=-0,63$ ;  $p=0,53$ ), Regresyon atama yöntemi sonucunda elde edilen veri setinde ortalamalar arasındaki fark anlamlı bulunmuştur ( $t=2,11$ ;  $p<0,05^*$ ). %10'luk, %15'lik ve %20'lik eksik veri setlerinde genel itibariyle EM atama yöntemi tam veri seti ile benzer sonuçlar vermiştir.

#### **4.5- 400 Birimlik Türetilmiş Veri Setleri**

Düşük ve yüksek korelasyonlara sahip 400 birimlik veri setlerine ilişkin analizler aşağıdaki bölümlerde verilmiştir.

#### 4.5.1- Düşük Korelasyonlu 400 Birimlik Veri Seti

400 birim içeren veri setine ilişkin değişkenler arasındaki korelasyonları gösteren Çizelge aşağıdadır. Çizelge 4.66'da görüleceği gibi değişkenler arasındaki korelasyon katsayıları oldukça düşüktür ( $r < 0.20$ ).

Çizelge 4.66- 400 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri

	X1	X2	X3	X4
X1	1	0.15 (P<0.01)	0.11 (P<0.05)	0.04 (P>0.05)
X2		1	0.05 (P>0.05)	0.03 (P>0.05)
X3			1	0.01 (P>0.05)
X4				1

Not: Çizelgede parantez içindeki değerler olasılık değerleridir.

Çizelge 4.67'de 400 birimden oluşan düşük korelasyonlara sahip tam veri setinde sırasıyla %5, %10, %15 ve %20 değer eksiltilerek elde edilen veri setlerindeki eksik olma sayı ve yüzdeleri verilmiştir. Çizelge incelendiğinde eksiltmelerde verilen yüzde değerlerine yaklaşık azaltmalar gerçekleşmiştir. Bu farklılaşmalar tam rasgelelik koşullarının sağlanması için yerine koyma yöntemi ile eksiltme yönteminin kullanılmasından kaynaklanmaktadır. Ancak öngörülen yüzde eksiltme değerlerine yaklaşmıştır.

Çizelge 4.67- 400 birim içeren düşük korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri

Değişken	Tam veri (n)	%5 Eksik		%10 Eksik		%15 Eksik		%20 Eksik	
		sayı	%	sayı	%	sayı	%	sayı	%
X1	400	20	5	38	9.5	58	14.5	72	18
X2	400	20	5	39	9.8	52	13	73	18.3
X3	400	20	5	36	9	58	14.5	70	17.5
X4	400	20	5	40	10	53	13.3	72	18
X5	400	0	0	0	0	0	0	0	0

Çizelge 4.68'de farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen ortalama ve standart sapma değerleri verilmiştir.

Çizelge 4.68- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri

		X1 $\bar{X} \pm S$	X2 $\bar{X} \pm S$	X3 $\bar{X} \pm S$	X4 $\bar{X} \pm S$
	<b>Tam Veri</b>	<b>0.104±0.987</b>	<b>-0.064±1.063</b>	<b>-0.051±0.981</b>	<b>-0.019±0.980</b>
%5 eksik	Listwise	0.074±0.979	-0.057±1.068	-0.058±0.989	-0.054±0.981
	Pairwise	0.090±0.985	-0.052±1.070	-0.049±0.974	-0.019±0.989
	Ortalama	0.090±0.960	-0.052±1.043	-0.049±0.949	-0.019±0.964
	EM	0.090±0.986	-0.052±1.070	-0.048±0.974	-0.019±0.989
	Regresyon	0.074±0.979	-0.057±1.068	-0.058±0.989	-0.054±0.98
%10 eksik	Listwise	0.111±1.030	-0.071±1.073	-0.036±1.009	-0.063±0.963
	Pairwise	0.111±1.001	-0.071±1.056	-0.014±0.985	-0.036±0.982
	Ortalama	0.111±0.952	-0.071±1.003	-0.014±0.939	-0.036±0.932
	EM	0.112±1.000	-0.074±1.056	-0.012±0.984	-0.035±0.982
	Regresyon	0.111±1.030	-0.071±1.073	-0.036±1.009	-0.063±0.963
%15 eksik	Listwise	0.141±0.936	-0.087±1.082	-0.034±0.967	-0.022±0.948
	Pairwise	0.142±0.968	-0.045±1.077	-0.034±0.975	-0.023±0.982
	Ortalama	0.142±0.895	-0.045±1.004	-0.034±0.902	-0.023±0.914
	EM	0.141±0.968	-0.043±1.077	-0.033±0.976	-0.023±0.982
	Regresyon	0.141±0.936	-0.087±1.082	-0.034±0.967	-0.022±0.948
%20 eksik	Listwise	0.166±1.035	-0.102±1.014	-0.006±0.989	-0.044±1.037
	Pairwise	0.099±1.003	-0.073±1.038	-0.049±0.991	-0.018±0.985
	Ortalama	0.099±0.908	-0.073±0.938	-0.049±0.900	-0.018±0.892
	EM	0.096±1.004	-0.084±1.039	-0.049±0.990	-0.018±0.985
	Regresyon	0.166±1.035	-0.102±1.014	-0.006±0.989	-0.044±1.037

Düşük korelasyonlu değişkenlere sahip 400 birimlik veri setlerinin analizleri sonucunda %5'lik, %10'luk ve %15'lik eksik veri setinde, genel olarak tüm atama yöntemlerinden elde edilen ortalama ve standart sapma değerleri, tam veri setinden elde edilen değerlere oldukça yakın bulunmuştur ( $P>0.05$ ). %20'lik eksik veri setinde ise Ortalama, Pairwise ve EM atama yöntemleri ile tam veri setinden elde edilen belirtici istatistikler arasında benzerlik gözlenmiştir ( $P>0.05$ ).

Çizelge 4.69'da tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen yeni veri setlerinden elde edilmiş regresyon katsayıları, bu katsayıların anlamlılık düzeyleri ve belirleyicilik katsayıları ( $R^2$ ) verilmiştir.

Çizelge 4.69- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları

		Sabit ( $b_0$ )	$b_1$	$b_2$	$b_3$	$R^2$
	<b>Tam Veri</b>	<b>0.118*</b>	<b>0.130**</b>	<b>0.106*</b>	<b>0.031</b>	<b>0.033</b>
%5 eksik	Listwise	0.086	0.102*	0.066	0.039	0.020
	Ortalama	0.100*	0.120**	0.066	0.025	0.023
	EM	0.101*	0.134**	0.073	0.026	0.029
	Regresyon	0.104*	0.160***	0.096	0.014	0.040
%10 eksik	Listwise	0.128*	0.156**	0.155*	0.016	0.052
	Ortalama	0.123**	0.134**	0.124*	0.025	0.036
	EM	0.126**	0.160**	0.138**	0.034	0.049
	Regresyon	0.130**	0.132**	0.144**	0.031	0.034
%15 eksik	Listwise	0.162**	0.160**	0.167**	0.052	0.071
	Ortalama	0.154**	0.121**	0.144**	0.047	0.045
	EM	0.156***	0.149**	0.205***	0.059	0.081
	Regresyon	0.169***	0.134**	0.160**	0.046	0.054
%20 eksik	Listwise	0.181*	0.138	0.197*	0.011	0.060
	Ortalama	0.115*	0.104*	0.156**	0.036	0.040
	EM	0.121**	0.171***	0.203***	0.044	0.082
	Regresyon	0.092	0.131**	0.166**	0.028	0.051

\*  $p < 0.05$       \*\*  $p < 0.01$       \*\*\*  $p < 0.001$

Düşük korelasyonlu değişkenlere sahip 400 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde yapılan regresyon analizlerinde ise EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_L^2=0.029$ ;  $R_T^2=0.033$ ). %10'luk eksik veri setlerinde ise Regresyon atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_R^2=0.034$ ;  $R_T^2=0.033$ ). %15'lik eksik veri setlerinde Regresyon atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_O^2=0.045$ ;  $R_T^2=0.033$ ). %20'lik eksik veri setlerinde ise tüm atama yöntemlerinden elde edilen belirleyicilik katsayıları, tam veri setinden elde edilen değere daha yakın çıkmıştır. Fakat sonuçlar, regresyon katsayılarının anlamlılıkları ve tam veri setine benzerliği olarak ele alındığında, EM ve Ortalama atama yöntemleri daha başarılı sonuçlar vermiştir.

Çizelge 4.70'de 400 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.70- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.151±1.050</b>	<b>0.056±0.920</b>	<b>0.96</b>	<b>398</b>	<b>0.34</b>
%5 eksik	Pairwise	0.127±1.049	0.054±0.921	0.73	378	0.47
	Ortalama	0.125±1.014	0.055±0.904	0.73	398	0.47
	EM	0.125±1.015	0.054±0.904	0.73	398	0.46
	Regresyon	0.115±1.057	0.059±0.920	0.57	398	0.57
%10 eksik	Pairwise	0.160±1.054	0.060±0.943	0.95	360	0.35
	Ortalama	0.156±1.008	0.066±0.892	0.95	398	0.35
	EM	0.161±1.010	0.062±0.893	1.04	398	0.30
	Regresyon	0.169±1.043	0.079±0.959	0.90	398	0.37
%15 eksik	Pairwise	0.217±1.030	0.070±0.900	1.40	340	0.16
	Ortalama	0.205±0.942	0.079±0.842	1.40	398	0.16
	EM	0.213±0.946	0.069±0.845	1.60	398	0.11
	Regresyon	0.216±1.011	0.092±0.914	1.29	398	0.20
%20 eksik	Pairwise	0.118±1.076	0.080±0.924	0.34	326	0.73
	Ortalama	0.115±0.983	0.084±0.828	0.34	398	0.73
	EM	0.124±0.988	0.068±0.834	0.61	398	0.54
	Regresyon	0.094±1.042	0.050±0.912	0.45	398	0.66

Çizelge 4.71'de 400 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X2 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.71- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.041±1.078</b>	<b>-0.087±1.049</b>	<b>0.43</b>	<b>398</b>	<b>0.67</b>
%5 eksik	Pairwise	-0.026±1.081	-0.080±1.061	0.49	378	0.63
	Ortalama	-0.027±1.062	-0.078±1.026	0.49	398	0.63
	EM	-0.023±1.062	-0.082±1.027	0.57	398	0.57
	Regresyon	-0.022±1.078	-0.094±1.064	0.67	398	0.50
%10 eksik	Pairwise	-0.062±1.052	-0.081±1.064	0.17	359	0.87
	Ortalama	-0.063±1.006	-0.079±1.003	0.17	398	0.87
	EM	-0.069±1.008	-0.079±1.004	0.10	398	0.92
	Regresyon	-0.035±1.045	-0.059±1.064	0.23	398	0.82
%15 eksik	Pairwise	-0.028±1.113	-0.060±1.045	0.28	346	0.78
	Ortalama	-0.031±1.023	-0.059±0.988	0.28	398	0.78
	EM	-0.025±1.025	-0.061±0.989	0.36	398	0.72
	Regresyon	-0.009±1.111	-0.050±1.027	0.38	398	0.70
%20 eksik	Pairwise	-0.065±1.072	-0.081±1.004	0.15	325	0.88
	Ortalama	-0.066±0.977	-0.080±0.900	0.15	398	0.88
	EM	-0.076±0.981	-0.091±0.904	0.17	398	0.87
	Regresyon	-0.081±1.047	-0.074±0.979	-0.06	398	0.95

Çizelge 4.72'de 400 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X3 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.72- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.042±0.950</b>	<b>-0.059±1.013</b>	<b>0.17</b>	<b>398</b>	<b>0.86</b>
%5 eksik	Pairwise	-0.047±0.947	-0.051±1.002	0.04	378	0.97
	Ortalama	-0.047±0.926	-0.051±0.974	0.04	398	0.97
	EM	-0.046±0.926	-0.051±0.974	0.05	398	0.96
	Regresyon	-0.053±0.946	-0.082±1.009	0.30	398	0.76
%10 eksik	Pairwise	0.004±0.961	-0.031±1.010	0.34	362	0.73
	Ortalama	0.026±0.917	-0.030±0.963	0.34	398	0.73
	EM	0.004±0.917	-0.029±0.963	0.35	398	0.73
	Regresyon	0.045±0.959	-0.022±1.010	0.68	398	0.49
%15 eksik	Pairwise	-0.020±0.932	-0.047±1.019	0.26	340	0.79
	Ortalama	-0.022±0.854	-0.045±0.949	0.26	398	0.79
	EM	-0.020±0.859	-0.046±0.951	0.29	398	0.77
	Regresyon	-0.039±0.898	-0.089±1.040	0.52	398	0.61
%20 eksik	Pairwise	-0.036±0.952	-0.062±1.030	0.24	328	0.81
	Ortalama	-0.039±0.849	-0.060±0.951	0.24	398	0.81
	EM	-0.042±0.852	-0.056±0.954	0.15	398	0.88
	Regresyon	-0.034±0.934	-0.084±1.017	0.51	398	0.61

Çizelge 4.73'de 400 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X4 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.73- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.035±1.041</b>	<b>-0.074±0.914</b>	<b>1.10</b>	<b>398</b>	<b>0.27</b>
%5 eksik	Pairwise	0.058±1.050	-0.095±0.920	1.51	378	0.13
	Ortalama	0.054±1.021	-0.092±0.899	1.51	398	0.13
	EM	0.054±1.021	-0.093±0.899	1.52	398	0.13
	Regresyon	0.054±1.042	-0.127±0.925	1.84	398	0.07
%10 eksik	Pairwise	0.026±1.059	-0.095±0.901	1.17	358	0.24
	Ortalama	0.018±0.994	-0.090±0.864	1.17	398	0.24
	EM	0.018±0.994	-0.089±0.864	1.15	398	0.25
	Regresyon	0.026±1.052	-0.095±0.884	1.25	398	0.21
%15 eksik	Pairwise	0.048±1.041	-0.095±0.915	1.36	345	0.18
	Ortalama	0.039±0.971	-0.085±0.851	1.36	398	0.17
	EM	0.040±0.972	-0.088±0.851	1.40	398	0.16
	Regresyon	0.074±1.008	-0.096±0.893	1.79	398	0.07
%20 eksik	Pairwise	0.032±1.033	-0.069±0.934	0.93	326	0.36
	Ortalama	0.023±0.941	-0.059±0.839	0.93	398	0.36
	EM	0.023±0.941	-0.059±0.840	0.92	398	0.36
	Regresyon	0.074±1.028	-0.055±0.918	1.32	398	0.19

Düşük korelasyonlu değişkenlere sahip 400 birimlik veri setlerinin analizleri sonucunda %5'lik, %10'luk ve %20'lik eksik veri setlerinde farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise genel itibariyle EM atama yöntemi

tam veri setine benzer sonuçlar vermiştir. %15'lik eksik veri setlerinde ise genel olarak tüm atama yöntemleri tam veri seti ile benzer sonuçlar vermiştir.

#### 4.5.2- Yüksek Korelasyonlu 400 Birimlik Veri Seti

400 birim içeren veri setine ilişkin değişkenler arasındaki korelasyonları gösteren Çizelge aşağıdadır. Çizelge 4.74'te görüleceği gibi değişkenler arasındaki korelasyon katsayıları oldukça yüksektir ( $r > 0.80$ ).

Çizelge 4.74- 400 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri

	X1	X2	X3	X4
X1	1	0.89 (P<0.001)	0.89 (P<0.001)	0.89 (P<0.001)
X2		1	0.90 (P<0.001)	0.91 (P<0.001)
X3			1	0.90 (P<0.001)
X4				1

Not: Çizelgede parantez içindeki değerler olasılık değerleridir.

Çizelge 4.75'te 400 birimden oluşan düşük korelasyonlara sahip tam veri setinde sırasıyla %5, %10, %15 ve %20 değer eksiltilerek elde edilen veri setlerindeki eksik olma sayı ve yüzdeleri verilmiştir. Çizelge incelendiğinde eksiltmelerde verilen yüzde değerlerine yaklaşık azaltmalar gerçekleşmiştir. Bu farklılaşmalar tam rasgelelik koşullarının sağlanması için yerine koyma yöntemi ile eksiltme yönteminin kullanılmasından kaynaklanmaktadır. Ancak öngörülen yüzde eksiltme değerlerine yaklaşmıştır.

Çizelge 4.75- 400 birim içeren yüksek korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri

Değişken	Tam veri (n)	%5 Eksik		%10 Eksik		%15 Eksik		%20 Eksik	
		sayı	%	sayı	%	sayı	%	sayı	%
X1	400	20	5	37	9.3	58	14.5	71	17.8
X2	400	20	5	39	9.8	57	14.3	77	19.3
X3	400	20	5	37	9.3	54	13.5	73	18.3
X4	400	20	5	37	9.3	54	13.5	72	18
X5	400	0	0	0	0	0	0	0	0

Çizelge 4.76'da farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen ortalama ve standart sapma değerleri verilmiştir.

Çizelge 4.76- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri

		X1 $\bar{X} \pm S$	X2 $\bar{X} \pm S$	X3 $\bar{X} \pm S$	X4 $\bar{X} \pm S$
	<b>Tam Veri</b>	<b>-0.008±0.994</b>	<b>0.006±0.13</b>	<b>-0.016±0.993</b>	<b>-0.021±0.996</b>
%5 eksik	Listwise	0.032±0.979	0.042±0.986	0.022±0.987	0.013±0.960
	Pairwise	-0.001±0.997	0.013±1.003	-0.013±1.001	-0.005±0.967
	Ortalama	-0.001±0.972	0.013±0.978	-0.013±0.976	-0.005±0.942
	EM	-0.008±0.996	0.011±1.013	-0.016±0.997	-0.024±0.976
	Regresyon	0.032±0.979	0.042±0.986	0.022±0.987	0.013±0.960
%10 eksik	Listwise	-0.015±1.015	0.007±1.049	-0.029±1.014	-0.058±0.979
	Pairwise	-0.017±0.998	-0.003±1.031	0.006±0.988	-0.048±0.950
	Ortalama	-0.017±0.951	-0.003±0.979	0.006±0.941	0.048±0.905
	EM	-0.012±0.989	0.010±1.019	-0.013±0.985	-0.036±0.954
	Regresyon	-0.015±1.015	0.007±1.049	-0.029±1.014	-0.058±0.979
%15 eksik	Listwise	0.012±1.028	0.007±1.044	0.008±1.013	-0.020±0.979
	Pairwise	0.010±0.992	-0.006±1.019	-0.008±0.989	-0.018±0.968
	Ortalama	0.010±0.917	-0.006±0.943	-0.008±0.920	-0.018±0.900
	EM	-0.015±0.993	0.006±1.014	-0.019±0.986	-0.024±0.953
	Regresyon	0.012±1.028	0.007±1.044	0.008±1.013	-0.020±0.979
%20 eksik	Listwise	0.043±1.043	0.084±1.037	0.063±1.010	0.016±0.960
	Pairwise	0.038±0.990	0.009±1.017	0.022±1.005	-0.033±0.960
	Ortalama	0.038±0.898	0.009±0.914	0.022±0.909	-0.033±0.869
	EM	-0.009±1.002	0.008±1.013	-0.010±0.998	-0.024±0.963
	Regresyon	0.043±1.043	0.084±1.037	0.063±1.010	0.016±0.960

Yüksek korelasyonlu değişkenlere sahip 400 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde, EM atama yöntemi tam veri setinden elde edilen belirtici istatistiklerle hemen hemen aynı sonucu vermiştir ( $P>0.05$ ). %10'luk eksik veri setlerinde Regresyon ve Listwise atama yöntemleri birbirleri arasında aynı sonuçları vermiştir ( $P>0.05$ ). Fakat burada EM atama yöntemi, tam veri setinden elde edilen belirtici istatistiklere benzer sonuçlar vermiştir ( $P>0.05$ ). %15'lik ve %20'lik eksik veri setlerinde, en iyi sonuçları EM atama yöntemi vermiştir. Ortalama ve standart sapma değerleri tam veri setine oldukça benzerdir ( $P>0.05$ ).

Çizelge 4.77'de tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen yeni veri setlerinden elde edilmiş regresyon katsayıları, bu katsayıların anlamlılık düzeyleri ve belirleyicilik katsayıları ( $R^2$ ) verilmiştir.



Çizelge 4.77- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları

		Sabit ( $b_0$ )	$b_1$	$b_2$	$b_3$	$R^2$
	<b>Tam Veri</b>	<b>0.003</b>	<b>0.262***</b>	<b>0.356***</b>	<b>0.341***</b>	<b>0.851</b>
%5 eksik	Listwise	0.009	0.255***	0.359***	0.341***	0.848
	Ortalama	0.003	0.231***	0.436***	0.277***	0.794
	EM	0.003	0.272***	0.366***	0.321***	0.860
	Regresyon	0.001	0.227***	0.391***	0.342***	0.852
%10 eksik	Listwise	0.013	0.263***	0.372***	0.325***	0.855
	Ortalama	-0.007	0.330***	0.368***	0.231***	0.757
	EM	0.002	0.275***	0.358***	0.331***	0.866
	Regresyon	0.001	0.329***	0.340***	0.295***	0.852
%15 eksik	Listwise	0.013	0.284***	0.369***	0.327***	0.867
	Ortalama	0.020	0.239***	0.334***	0.335***	0.698
	EM	-0.001	0.274***	0.352***	0.352***	0.882
	Regresyon	0.009	0.259***	0.390***	0.313***	0.855
%20 eksik	Listwise	-0.010	0.358***	0.351***	0.284**	0.859
	Ortalama	0.033	0.351***	0.337***	0.193***	0.649
	EM	0.00001	0.316***	0.321***	0.341***	0.885
	Regresyon	0.0001	0.445***	0.349***	0.165**	0.854

\*  $p < 0.05$       \*\*  $p < 0.01$       \*\*\*  $p < 0.001$

Yüksek korelasyonlu değişkenlere sahip 400 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setlerinde yapılan regresyon analizlerinde ise Regresyon ve Listwise atama yöntemlerinden elde edilen belirleyicilik katsayıları, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_R^2=0.852$ ;  $R_L^2=0.848$   $R_T^2=0.851$ ). Burada EM atama yöntemiyle elde edilen veri setinin belirleyicilik katsayısı da regresyon katsayısına yakın bulunmuştur ( $R_{EM}^2=0.860$ ). %10'luk eksik veri setlerinde Regresyon atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_R^2=0.852$ ;  $R_T^2=0.851$ ). EM atama yöntemiyle elde edilen veri setinin belirleyicilik katsayısı ise tam veri setinden elde edilen katsayıya oldukça yakın çıkmıştır ( $R_{EM}^2=0.866$ ). %15'lik eksik veri setlerinde Regresyon atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_R^2=0.855$ ;  $R_T^2=0.851$ ). EM atama yöntemiyle elde edilen veri setinin belirleyicilik katsayısı tam veri setine yakın bulunmuştur ( $R_{EM}^2=0.882$ ). %20'lik eksik veri setlerinde ise Regresyon ve Listwise atama yöntemlerinden elde edilen

belirleyicilik katsayıları, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_R^2=0.854$ ;  $R_{EM}^2=0.885$ ;  $R_T^2=0.851$ ). Fakat sonuçlar, regresyon katsayılarının anlamlılıkları ve tam veri setine benzerliği olarak ele alındığında EM ve Ortalama atama yöntemi daha başarılı sonuçlar vermiştir.

Çizelge 4.78’de 400 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.78- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.018±0.905</b>	<b>0.001±1.071</b>	<b>-0.19</b>	<b>398</b>	<b>0.85</b>
%5 eksik	Pairwise	-0.018±0.912	0.015±1.073	-0.33	378	0.75
	Ortalama	-0.018±0.895	0.014±1.039	-0.33	398	0.75
	EM	-0.024±0.904	0.007±1.068	-0.31	398	0.76
	Regresyon	-0.028±0.903	0.003±1.083	-0.31	398	0.76
%10 eksik	Pairwise	-0.026±0.903	-0.008±1.085	-0.17	361	0.87
	Ortalama	-0.025±0.874	-0.088±1.018	-0.17	398	0.87
	EM	-0.016±0.896	-0.009±1.057	-0.08	398	0.94
	Regresyon	-0.010±0.893	-0.013±1.069	0.04	398	0.97
%15 eksik	Pairwise	-0.023±0.897	0.041±1.076	-0.59	340	0.55
	Ortalama	-0.019±0.836	0.036±0.987	-0.59	398	0.55
	EM	-0.027±0.887	-0.004±1.063	-0.23	398	0.82
	Regresyon	-0.013±0.881	0.009±1.069	-0.22	398	0.83
%20 eksik	Pairwise	0.032±0.897	0.043±1.071	-0.10	327	0.92
	Ortalama	0.033±0.815	0.042±0.968	-0.10	398	0.92
	EM	-0.019±0.895	0.0001±1.068	-0.19	398	0.85
	Regresyon	-0.017±0.919	0.0001±1.068	-0.17	398	0.86

Çizelge 4.79’da 400 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X2 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.79- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.005±0.924</b>	<b>0.006±1.091</b>	<b>-0.01</b>	<b>398</b>	<b>0.99</b>
%5 eksik	Pairwise	0.013±0.924	0.012±1.075	0.01	378	0.99
	Ortalama	0.013±0.907	0.012±1.041	0.01	398	0.99
	EM	0.014±0.920	0.007±1.087	0.07	398	0.95
	Regresyon	0.011±0.925	0.006±1.086	0.05	398	0.96
%10 eksik	Pairwise	-0.014±0.932	0.007±1.116	-0.19	359	0.85
	Ortalama	-0.013±0.885	0.063±1.061	-0.19	398	0.85
	EM	0.009±0.929	0.010±1.086	-0.01	398	0.99
	Regresyon	-0.005±0.933	0.014±1.088	-0.18	398	0.86
%15 eksik	Pairwise	-0.012±0.928	0.001±1.101	-0.12	341	0.91
	Ortalama	-0.011±0.867	-0.028±1.010	-0.12	398	0.91
	EM	0.009±0.927	0.003±1.071	0.06	398	0.95
	Regresyon	0.017±0.930	0.005±1.073	0.12	398	0.91
%20 eksik	Pairwise	0.048±0.926	-0.024±1.091	0.63	321	0.53
	Ortalama	0.039±0.817	-0.018±0.995	0.63	398	0.53
	EM	0.007±0.913	0.010±1.072	-0.03	398	0.98
	Regresyon	-0.011±0.937	-0.009±1.076	-0.02	398	0.99

Çizelge 4.80'de 400 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X3 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.80- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.042±0.931</b>	<b>0.008±1.049</b>	<b>-0.51</b>	<b>398</b>	<b>0.61</b>
%5 eksik	Pairwise	-0.045±0.929	0.017±1.064	-0.60	378	0.55
	Ortalama	-0.044±0.907	0.015±1.036	-0.60	398	0.55
	EM	-0.047±0.928	0.012±1.050	-0.59	398	0.56
	Regresyon	-0.054±0.932	0.017±1.057	-0.71	398	0.48
%10 eksik	Pairwise	-0.027±0.941	0.036±1.031	-0.61	361	0.54
	Ortalama	-0.024±0.893	0.034±0.985	-0.61	398	0.54
	EM	-0.045±0.916	0.016±1.034	-0.61	398	0.54
	Regresyon	-0.042±0.918	0.016±1.036	-0.59	398	0.56
%15 eksik	Pairwise	-0.042±0.919	0.023±1.051	-0.61	344	0.55
	Ortalama	-0.037±0.854	0.019±0.978	-0.61	398	0.55
	EM	-0.042±0.914	0.002±1.030	-0.45	398	0.65
	Regresyon	-0.038±0.924	0.0001±1.038	-0.39	398	0.70
%20 eksik	Pairwise	0.001±0.944	0.041±1.058	-0.36	325	0.72
	Ortalama	0.005±0.844	0.037±0.965	-0.36	398	0.72
	EM	-0.032±0.920	0.010±1.039	-0.43	398	0.67
	Regresyon	-0.023±0.927	0.029±1.061	-0.52	398	0.60

Çizelge 4.81'de 400 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X4 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.81- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.034±0.885</b>	<b>-0.009±1.037</b>	<b>-0.26</b>	<b>398</b>	<b>0.80</b>
%5 eksik	Pairwise	-0.040±0.894	0.028±1.031	-0.69	378	0.49
	Ortalama	-0.039±0.875	0.026±1.001	-0.69	398	0.49
	EM	-0.039±0.892	-0.011±1.043	-0.29	398	0.77
	Regresyon	-0.037±0.900	-0.013±1.046	-0.24	398	0.81
%10 eksik	Pairwise	-0.038±0.860	-0.056±1.027	0.18	361	0.86
	Ortalama	-0.039±0.818	-0.055±0.979	0.18	398	0.86
	EM	-0.038±0.870	-0.033±1.017	-0.05	398	0.96
	Regresyon	-0.040±0.861	-0.033±1.012	-0.08	398	0.94
%15 eksik	Pairwise	-0.051±0.871	0.012±1.050	-0.60	344	0.55
	Ortalama	-0.046±0.806	0.008±0.980	-0.60	398	0.55
	EM	-0.041±0.860	-0.010±1.017	-0.32	398	0.75
	Regresyon	-0.035±0.859	-0.002±1.015	-0.35	398	0.73
%20 eksik	Pairwise	-0.079±0.854	0.007±1.045	-0.81	326	0.42
	Ortalama	-0.070±0.767	0.0001±0.953	-0.81	398	0.42
	EM	-0.048±0.861	-0.002±1.026	-0.48	398	0.63
	Regresyon	-0.061±0.872	-0.002±1.034	-0.61	398	0.54

Yüksek korelasyonlu değişkenlere sahip 400 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setinde farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise genel itibariyle EM atama yöntemi tam veri seti ile benzer sonuçlar verse de diğer atama yöntemleri ile elde edilen sonuçlarda benzerlik göstermektedir. %10'luk eksik veri genel itibariyle Pairwise, Ortalama ve EM atama yöntemleri benzer sonuçlar vermiştir. Regresyon atama yönteminden tutarsız sonuçlar elde edilmiştir. %15'lik ve %20'lik eksik veri setlerinde ortalamalar arasındaki farkların saptanmasında ise genel itibariyle EM atama yöntemi tam veri setinden elde edilen değerlere benzer sonuçlar vermiştir.

#### **4.6- 500 Birimlik Türetilmiş Veri Setleri**

Düşük ve yüksek korelasyonlara sahip 500 birimlik veri setlerine ilişkin analizler aşağıdaki bölümlerde verilmiştir.

##### **4.6.1- Düşük Korelasyonlu 500 Birimlik Veri Seti**

500 birim içeren veri setine ilişkin değişkenler arasındaki korelasyonları gösteren Çizelge aşağıdadır. Çizelge 4.82'de görüleceği gibi değişkenler arasındaki korelasyon katsayıları oldukça düşüktür ( $r < 0.40$ ).

Çizelge 4.82- 500 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri

	X1	X2	X3	X4
X1	1	0.14 (P<0.01)	0.16(P<0.001)	0.15 (P<0.01)
X2		1	0.11 (P<0.05)	0.10 (P<0.05)
X3			1	0.08 (P>0.05)
X4				1

Not: Çizelgede parantez içindeki değerler olasılık değerleridir.

Çizelge 4.83'te 500 birimden oluşan düşük korelasyonlara sahip tam veri setinde sırasıyla %5, %10, %15 ve %20 değer eksiltilecek elde edilen veri setlerindeki eksik olma sayı ve yüzdeleri verilmiştir. Çizelge incelendiğinde eksiltmelerde verilen yüzde değerlerine yaklaşık azaltmalar gerçekleşmiştir. Bu farklılaşmalar tam rasgelelik koşullarının sağlanması için yerine koyma yöntemi ile eksiltme yönteminin kullanılmasından kaynaklanmaktadır. Ancak öngörülen yüzde eksiltme değerlerine yaklaşılmıştır.

Çizelge 4.83- 500 birim içeren düşük korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri

Değişken	Tam veri (n)	%5 Eksik		%10 Eksik		%15 Eksik		%20 Eksik	
		sayı	%	sayı	%	sayı	%	sayı	%
X1	500	25	5	50	10	70	14	90	18
X2	500	25	5	47	9.4	73	14.6	90	18
X3	500	25	5	48	9.6	67	13.4	90	18
X4	500	25	5	47	9.4	72	14.4	91	18.2
X5	500	0	0	0	0	0	0	0	0

Çizelge 4.84'te farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen ortalama ve standart sapma değerleri verilmiştir.

Düşük korelasyonlu değişkenlere sahip 500 birimlik veri setlerinin analizleri sonucunda %5'lik ve %15'lik eksik veri setlerinde, EM atama yöntemi tam veri setinden elde edilen belirtici istatistiklere oldukça yakın sonuçlar vermiştir (P>0.05). %10'luk eksik veri setlerinde, tüm atama yöntemleri benzer sonuçlar vermişlerdir (P>0.05). %20'luk eksik veri setinde de tüm atama yöntemleri benzer sonuçlar vermiş olsalarda tam veri setine en yakın ortalamalar Ortalama, Pairwise ve EM atama yöntemlerinden elde edilmiştir (P>0.05).

Çizelge 4.84- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri

		X1 $\bar{X} \pm S$	X2 $\bar{X} \pm S$	X3 $\bar{X} \pm S$	X4 $\bar{X} \pm S$
	<b>Tam Veri</b>	<b>0.029±1.016</b>	<b>0.020±1.049</b>	<b>0.013±1.007</b>	<b>0.051±0.951</b>
%5 eksik	Listwise	0.082±1.009	-0.003±1.053	0.038±1.019	0.032±0.949
	Pairwise	0.040±1.012	0.031±1.057	0.025±1.011	0.040±0.944
	Ortalama	0.040±0.986	0.031±1.030	0.025±0.985	0.040±0.920
	EM	0.039±1.012	0.032±1.057	0.025±1.011	0.038±0.943
	Regresyon	0.082±1.009	-0.003±1.053	0.038±1.019	0.032±0.949
%10 eksik	Listwise	0.025±1.024	-0.054±1.061	-0.008±1.001	0.032±0.990
	Pairwise	0.017±1.025	0.022±1.062	0.016±1.011	0.069±0.957
	Ortalama	0.017±0.972	0.022±1.010	0.016±0.961	0.069±0.911
	EM	0.024±1.025	0.025±1.062	0.016±1.010	0.069±0.957
	Regresyon	0.025±1.024	-0.054±1.061	-0.008±1.001	0.032±0.990
%15 eksik	Listwise	-0.001±1.014	-0.019±1.060	-0.036±0.988	0.075±0.912
	Pairwise	0.006±1.016	0.015±1.042	-0.001±1.010	0.070±0.934
	Ortalama	0.006±0.942	0.015±0.963	-0.001±0.940	0.070±0.864
	EM	0.005±1.017	0.016±1.042	-0.002±1.010	0.071±0.934
	Regresyon	-0.001±1.014	-0.019±1.060	-0.036±0.988	0.075±0.912
%20 eksik	Listwise	-0.032±0.956	-0.008±1.034	-0.087±1.028	0.012±0.947
	Pairwise	0.002±0.994	-0.043±1.053	0.019±1.009	0.027±0.958
	Ortalama	0.002±0.900	-0.043±0.954	0.019±0.914	0.027±0.866
	EM	0.004±0.994	-0.042±1.052	0.024±1.010	0.025±0.958
	Regresyon	-0.032±0.956	-0.008±1.034	-0.087±1.028	0.012±0.947

Çizelge 4.85'te tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen yeni veri setlerinden elde edilmiş regresyon katsayıları, bu katsayıların anlamlılık düzeyleri ve belirleyicilik katsayıları ( $R^2$ ) verilmiştir.

Çizelge 4.85- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları

		Sabit ( $b_0$ )	$b_1$	$b_2$	$b_3$	$R^2$
	<b>Tam Veri</b>	<b>0.019</b>	<b>0.104*</b>	<b>0.139**</b>	<b>0.140**</b>	<b>0.057</b>
%5 eksik	Listwise	0.073	0.093*	0.090	0.194***	0.058
	Ortalama	0.029	0.073	0.102*	0.157**	0.043
	EM	0.027	0.079	0.110*	0.173***	0.047
	Regresyon	0.022	0.076	0.085	0.175***	0.048
%10 eksik	Listwise	0.030	0.120*	0.155**	0.106	0.051
	Ortalama	0.004	0.098*	0.128**	0.127**	0.048
	EM	0.089	0.128**	0.147**	0.144**	0.064
	Regresyon	0.021	0.125**	0.140**	0.144**	0.062
%15 eksik	Listwise	-0.004	0.087	0.163**	0.149*	0.049
	Ortalama	-0.004	0.062	0.112*	0.128**	0.035
	EM	-0.008	0.086*	0.145**	0.162**	0.056
	Regresyon	0.002	0.066	0.149**	0.192***	0.064
%20 eksik	Listwise	-0.020	0.100	0.157*	0.049	0.037
	Ortalama	0.001	0.102*	0.128**	0.095*	0.043
	EM	0.002	0.130**	0.184**	0.133**	0.084
	Regresyon	-0.010	0.104*	0.167***	0.087	0.049

\*  $p < 0.05$       \*\*  $p < 0.01$       \*\*\*  $p < 0.001$

Düşük korelasyonlu değişkenlere sahip 500 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri setlerinde yapılan regresyon analizlerinde ise Listwise, EM ve Regresyon atama yöntemlerinden elde edilen belirleyicilik katsayıları, tam veri setinden elde edilen değere daha yakın çıkmıştır ( $R_L^2=0.058$ ;  $R_{EM}^2=0.047$ ;  $R_R^2=0.048$ ;  $R_T^2=0.057$ ). Fakat sonuçlar, regresyon katsayılarının anlamlılıkları ve tam veri setine benzerliği olarak ele alındığında EM ve Ortalama atama yöntemleri daha başarılı sonuçlar vermiştir.

%10'luk eksik veri setlerinde, Regresyon, Listwise ve EM atama yöntemlerinden elde edilen belirleyicilik katsayıları ve regresyon katsayılarının anlamlılık düzeyleri birbirlerine oldukça yakın sonuçlar vermiştir ( $R_R^2=0.062$ ;  $R_L^2=0.051$ ;  $R_{EM}^2=0.064$ ;  $R_T^2=0.057$ ). Burada regresyon katsayılarının anlamlılık düzeyleri göz önüne alındığında ise Regresyon ve EM atama yöntemleri ile yapılan tahminler daha başarılıdır.

%15'lik eksik veri setlerinde EM atama yönteminden elde edilen belirleyicilik katsayısı, tam veri setinden elde edilen değere oldukça yakın çıkmıştır ( $R_{EM}^2=0.056$ ;  $R_T^2=0.057$ ). Regresyon katsayılarının anlamlılıkları ve tam veri setine benzerliği olarak ele alındığında da EM atama yöntemi daha başarılı sonuçlar vermiştir.

%20'luk eksik veri setlerinde ise Regresyon, Ortalama, Listwise ve EM atama yöntemlerinden elde edilen belirleyicilik katsayıları birbirlerine oldukça yakın sonuçlar vermiştir ( $R_R^2=0.049$ ;  $R_O^2=0.043$ ;  $R_L^2=0.037$ ;  $R_{EM}^2=0.084$ ;  $R_T^2=0.057$ ). Bu analizlerde regresyon katsayıları ve anlamlılık düzeyleri göz önüne alındığında ise EM atama yöntemi ile yapılan tahminler daha başarılıdır.

Çizelge 4.86'da 500 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.86- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.053±0.944</b>	<b>0.009±1.075</b>	<b>0.48</b>	<b>498</b>	<b>0.63</b>
%5 eksik	Pairwise	0.066±0.947	0.019±1.065	0.51	473	0.61
	Ortalama	0.064±0.922	0.020±1.039	0.51	498	0.61
	EM	0.061±0.924	0.020±1.040	0.47	498	0.64
	Regresyon	0.063±0.947	0.008±1.056	0.61	498	0.55
%10 eksik	Pairwise	0.064±0.948	-0.025±1.090	0.92	448	0.36
	Ortalama	0.060±0.913	-0.020±1.021	0.92	498	0.36
	EM	0.061±0.914	-0.007±1.025	0.78	498	0.44
	Regresyon	0.062±0.962	0.005±1.054	0.85	498	0.40
%15 eksik	Pairwise	0.037±0.937	-0.021±1.082	0.59	428	0.56
	Ortalama	0.033±0.873	-0.017±0.998	0.59	498	0.56
	EM	0.025±0.878	-0.013±1.002	0.45	498	0.66
	Regresyon	0.051±0.966	-0.020±1.077	0.77	498	0.44
%20 eksik	Pairwise	0.053±0.939	-0.045±1.041	0.99	408	0.32
	Ortalama	0.045±0.863	-0.035±0.930	0.99	498	0.32
	EM	0.047±0.868	-0.032±0.936	0.98	498	0.33
	Regresyon	0.045±0.959	-0.049±1.039	1.04	498	0.30

Çizelge 4.87’de 500 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X2 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.87- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.023±1.072</b>	<b>0.056±1.030</b>	<b>-0.85</b>	<b>498</b>	<b>0.40</b>
%5 eksik	Pairwise	-0.024±1.086	0.078±1.031	-1.05	473	0.30
	Ortalama	-0.021±1.057	0.076±1.006	-1.05	498	0.30
	EM	-0.017±1.058	0.074±1.006	-0.99	498	0.32
	Regresyon	-0.022±1.086	0.047±1.037	-0.73	498	0.47
%10 eksik	Pairwise	-0.036±1.095	0.073±1.031	-1.09	451	0.28
	Ortalama	-0.031±1.042	0.068±0.982	-1.09	498	0.28
	EM	-0.035±1.044	0.076±0.984	-1.22	498	0.22
	Regresyon	-0.018±1.085	0.069±1.019	-0.92	498	0.36
%15 eksik	Pairwise	-0.026±1.067	0.051±1.020	-0.76	425	0.45
	Ortalama	-0.021±0.995	0.046±0.935	-0.77	498	0.44
	EM	-0.019±0.998	0.047±0.938	-0.76	498	0.45
	Regresyon	0.014±1.066	0.044±0.997	-0.33	498	0.75
%20 eksik	Pairwise	-0.101±1.083	0.009±1.025	-1.06	408	0.29
	Ortalama	-0.092±0.992	0.001±0.919	-1.06	498	0.29
	EM	-0.090±0.995	0.0001±0.921	-1.05	498	0.29
	Regresyon	-0.081±1.056	-0.008±1.033	-0.78	498	0.44

Çizelge 4.88’de 500 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X3 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.



Çizelge 4.88- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.009±0.995</b>	<b>0.032±1.019</b>	<b>-0.45</b>	<b>498</b>	<b>0.65</b>
%5 eksik	Pairwise	-0.008±0.993	0.054±1.028	-0.66	473	0.51
	Ortalama	-0.007±0.973	0.052±0.996	-0.66	498	0.51
	EM	-0.007±0.974	0.053±0.997	-0.67	498	0.50
	Regresyon	0.029±1.027	0.049±1.013	-0.22	498	0.83
%10 eksik	Pairwise	-0.014±1.009	0.041±1.013	-0.58	450	0.56
	Ortalama	-0.011±0.953	0.039±0.968	-0.58	498	0.57
	EM	-0.010±0.954	0.038±0.970	-0.56	498	0.58
	Regresyon	0.012±1.001	0.072±1.014	-0.66	498	0.51
%15 eksik	Pairwise	0.005±0.993	-0.007±1.026	0.13	431	0.90
	Ortalama	0.004±0.922	-0.006±0.957	0.13	498	0.90
	EM	0.005±0.923	-0.009±0.959	0.17	498	0.86
	Regresyon	-0.007±0.984	-0.023±1.011	0.18	498	0.86
%20 eksik	Pairwise	0.001±1.002	0.035±1.017	-0.34	408	0.74
	Ortalama	0.005±0.906	0.032±0.922	-0.34	498	0.74
	EM	0.008±0.911	0.037±0.926	-0.36	498	0.72
	Regresyon	0.034±1.017	0.051±1.007	-0.19	498	0.85

Çizelge 4.89’da 500 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X4 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.89- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>0.026±0.937</b>	<b>0.072±0.968</b>	<b>-0.54</b>	<b>498</b>	<b>0.59</b>
%5 eksik	Pairwise	0.037±0.923	0.042±0.963	-0.05	473	0.96
	Ortalama	0.037±0.900	0.041±0.938	-0.05	498	0.96
	EM	0.064±0.901	0.040±0.938	-0.07	498	0.94
	Regresyon	0.021±0.920	0.056±0.968	-0.41	498	0.68
%10 eksik	Pairwise	0.040±0.936	0.094±0.977	-0.60	451	0.55
	Ortalama	0.043±0.894	0.092±0.926	-0.60	498	0.55
	EM	0.042±0.895	0.092±0.928	-0.61	498	0.54
	Regresyon	0.042±0.923	0.095±0.972	-0.62	498	0.54
%15 eksik	Pairwise	0.036±0.899	0.101±0.964	-0.72	426	0.47
	Ortalama	0.040±0.837	0.096±0.887	-0.72	498	0.47
	EM	0.038±0.840	0.100±0.889	-0.79	498	0.43
	Regresyon	0.019±0.907	0.097±0.963	-0.93	498	0.35
%20 eksik	Pairwise	0.020±0.929	0.033±0.985	-0.14	407	0.89
	Ortalama	0.021±0.847	0.032±0.884	-0.14	498	0.89
	EM	0.019±0.848	0.030±0.886	-0.14	498	0.89
	Regresyon	0.023±0.908	0.043±0.982	-0.23	498	0.82

Düşük korelasyonlu değişkenlere sahip 500 birimlik veri setlerinin analizleri sonucunda %5’lik, %10’luk, %15’lik ve %20’lik eksik veri setlerinin tamamında farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında genel itibariyle EM atama

yöntemi, tam veri seti ile oldukça benzer sonuçlar vermiştir. %10'luk ve %20'lik veri setlerinde genel itibariyle hemen hemen tüm atama yöntemlerinden benzer sonuçlar elde edilmiştir.

#### 4.6.2- Yüksek Korelasyonlu 500 Birimlik Veri Seti

500 birim içeren veri setine ilişkin değişkenler arasındaki korelasyonları gösteren Çizelge aşağıdadır. Çizelge 4.90'da görüleceği gibi değişkenler arasındaki korelasyon katsayıları oldukça yüksektir ( $r > 0.80$ ).

Çizelge 4.90- 500 birimlik veri setinde korelasyon katsayıları ve anlamlılık düzeyleri

	X1	X2	X3	X4
X1	1	0.90 (P<0.001)	0.90 (P<0.001)	0.91 (P<0.001)
X2		1	0.90 (P<0.001)	0.91 (P<0.001)
X3			1	0.91 (P<0.001)
X4				1

Not: Çizelgede parantez içindeki değerler olasılık değerleridir.

Çizelge 4.91'de 500 birimden oluşan düşük korelasyonlara sahip tam veri setinde sırasıyla %5, %10, %15 ve %20 değer eksiltilerek elde edilen veri setlerindeki eksik olma sayı ve yüzdeleri verilmiştir. Çizelge incelendiğinde eksiltmelerde verilen yüzde değerlerine yaklaşık azaltmalar gerçekleşmiştir. Bu farklılaşmalar tam rasgelelik koşullarının sağlanması için yerine koyma yöntemi ile eksiltme yönteminin kullanılmasından kaynaklanmaktadır. Ancak öngörülen yüzde eksiltme değerlerine yaklaşmıştır.

Çizelge 4.91- 500 birim içeren yüksek korelasyonlu değişkenlere sahip veri setinde, değişkenlerin eksik olma sayı ve yüzdeleri

Değişken	Tam veri (n)	%5 Eksik		%10 Eksik		%15 Eksik		%20 Eksik	
		sayı	%	sayı	%	sayı	%	sayı	%
X1	500	25	5	48	9.6	69	13.8	89	17.8
X2	500	25	5	47	9.4	71	14.2	93	18.6
X3	500	25	5	47	9.4	73	14.6	90	18
X4	500	25	5	48	9.6	70	14	92	18.4
X5	500	0	0	0	0	0	0	0	0

Çizelge 4.92’de farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen ortalama ve standart sapma değerleri verilmiştir.

Çizelge 4.92- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait ortalama ve standart sapma değerleri

		X1 $\bar{X} \pm S$	X2 $\bar{X} \pm S$	X3 $\bar{X} \pm S$	X4 $\bar{X} \pm S$
	<b>Tam Veri</b>	<b>-0.001±1.026</b>	<b>0.005±1.043</b>	<b>0.001±1.026</b>	<b>-0.028±1.020</b>
%5 eksik	Listwise	-0.013±1.032	-0.022±1.052	0.0003±1.028	-0.044±1.035
	Pairwise	0.008±1.023	-0.002±1.055	-0.002±1.021	-0.034±1.026
	Ortalama	0.008±0.997	-0.002±1.028	-0.002±0.995	-0.034±1.000
	EM	0.004±1.024	0.001±1.048	0.007±1.020	-0.029±1.022
	Regresyon	-0.013±1.032	-0.022±1.052	0.0003 ±1.028	-0.044±1.035
%10 eksik	Listwise	0.008±1.030	0.029±1.032	0.030±1.025	-0.018±1.014
	Pairwise	-0.007±1.032	0.014±1.035	0.011±1.031	-0.040±1.010
	Ortalama	-0.007±0.981	0.014±0.985	0.011±0.982	-0.040±0.960
	EM	-0.010±1.027	0.010±1.036	-0.001±1.030	-0.040±1.022
	Regresyon	0.008±1.030	0.029±1.032	0.030±1.025	-0.018±1.014
%15 eksik	Listwise	0.013±1.064	-0.004±1.107	0.001±1.084	-0.024±1.058
	Pairwise	-0.008±1.040	-0.009±1.035	0.001±1.045	-0.029±1.032
	Ortalama	-0.008±0.966	-0.009±0.958	0.001±0.965	-0.029±0.957
	EM	-0.002±1.027	-0.004±1.037	0.004±1.034	-0.031±1.016
	Regresyon	0.013±1.064	-0.004±1.107	0.001±1.084	-0.024±1.058
%20 eksik	Listwise	-0.008±1.007	-0.013±1.018	-0.043±1.003	-0.050±0.984
	Pairwise	-0.008±1.042	-0.024±1.045	-0.011±1.022	-0.026±1.011
	Ortalama	-0.008±0.945	-0.024±0.942	-0.011±0.925	-0.026±0.913
	EM	0.011±1.038	0.011±1.055	-0.015±1.038	-0.027±1.015
	Regresyon	-0.008±1.007	-0.013±1.018	-0.043±1.003	-0.050±0.984

Yüksek korelasyonlu değişkenlere sahip 500 birimlik veri setlerinin analizleri sonucunda %5’lik eksik veri setlerinde, X1 ve X4 değişkenlerinde EM atama yöntemi ( $P>0.05$ ), X2 ve X3 değişkenlerinde ise Pairwise ve Ortalama atama yöntemlerinden elde edilen belirtici istatistikler tam veri setinden elde edilen belirtici istatistiklere oldukça yakın sonuçlar vermiştir ( $P>0.05$ ). %10’luk ve %15’lik eksik veri setlerinde ise 4 değişken göz önüne alındığında, EM atama yöntemi, tam veri setinden elde edilen belirtici istatistiklere oldukça yakın sonuçlar vermiştir ( $P>0.05$ ). %20’lik eksik veri setlerinde ise hemen hemen tüm atama yöntemleri ile elde edilen belirtici istatistikler birbirlerine yakın sonuçlar vermişlerdir ( $P>0.05$ ).

Çizelge 4.93’te tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinden elde edilen yeni veri setlerinden elde edilmiş regresyon katsayıları, bu katsayıların anlamlılık düzeyleri ve belirleyicilik katsayıları ( $R^2$ ) verilmiştir.

Çizelge 4.93- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerine ait regresyon analizi sonuçları

		Sabit ( $b_0$ )	$b_1$	$b_2$	$b_3$	$R^2$
	<b>Tam Veri</b>	<b>0.008</b>	<b>0.299***</b>	<b>0.294***</b>	<b>0.365***</b>	<b>0.866</b>
%5 eksik	Listwise	0.009	0.297***	0.303***	0.353***	0.865
	Ortalama	0.022	0.314***	0.236***	0.375***	0.801
	EM	0.013	0.306***	0.287***	0.363***	0.870
	Regresyon	0.019	0.321***	0.258***	0.374***	0.861
%10 eksik	Listwise	-0.003	0.302***	0.290***	0.372***	0.861
	Ortalama	0.0002	0.277***	0.296***	0.363***	0.760
	EM	0.003	0.285***	0.287***	0.389***	0.877
	Regresyon	0.009	0.296***	0.283***	0.379***	0.867
%15 eksik	Listwise	0.023	0.306***	0.289***	0.354***	0.874
	Ortalama	0.004	0.311***	0.292***	0.325***	0.857
	EM	0.007	0.297***	0.365***	0.305***	0.888
	Regresyon	-0.010	0.278***	0.400***	0.276***	0.849
%20 eksik	Listwise	0.027	0.307***	0.283***	0.377***	0.859
	Ortalama	0.011	0.370***	0.205***	0.334***	0.652
	EM	0.022	0.308***	0.274***	0.396***	0.901
	Regresyon	0.036*	0.341***	0.237***	0.387***	0.871

\*  $p < 0.05$       \*\*  $p < 0.01$       \*\*\*  $p < 0.001$

Yüksek korelasyonlu değişkenlere sahip 500 birimlik veri setlerinin analizleri sonucunda %5'lik eksik veri yapılan regresyon analizlerinde ise Regresyon, Listwise ve EM atama yöntemlerinden elde edilen belirleyicilik katsayıları ve regresyon katsayılarının anlamlılık düzeyleri birbirlerine oldukça yakın sonuçlar vermiştir ( $R_R^2=0.861$ ;  $R_L^2=0.865$ ;  $R_{EM}^2=0.870$ ;  $R_T^2=0.866$ ). %10'luk eksik veri setlerinde Regresyon, Listwise ve EM atama yöntemlerinden elde edilen belirleyicilik katsayıları ve regresyon katsayılarının anlamlılık düzeyleri birbirlerine oldukça yakın sonuçlar vermiştir ( $R_R^2=0.867$ ;  $R_L^2=0.861$ ;  $R_{EM}^2=0.877$ ;  $R_T^2=0.866$ ). %15'lik eksik veri setlerinde, Regresyon, Ortalama, Listwise ve EM atama yöntemlerinden elde edilen belirleyicilik katsayıları ve regresyon katsayılarının anlamlılık düzeyleri birbirlerine oldukça yakın sonuçlar vermiştir ( $R_R^2=0.849$ ;  $R_L^2=0.874$ ;  $R_{EM}^2=0.888$ ;  $R_T^2=0.866$ ). %20'luk eksik veri setlerinde yapılan regresyon analizlerinde ise Regresyon, Listwise ve EM atama yöntemlerinden elde edilen belirleyicilik katsayıları ve regresyon katsayılarının anlamlılık düzeyleri birbirlerine oldukça yakın sonuçlar vermiştir ( $R_R^2=0.871$ ;  $R_L^2=0.859$ ;  $R_{EM}^2=0.901$ ;  $R_T^2=0.866$ ).

Çizelge 4.94'te 500 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X1 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.94- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X1 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.047±0.996</b>	<b>0.049±1.056</b>	<b>-1.06</b>	<b>498</b>	<b>0.29</b>
%5 eksik	Pairwise	-0.027±0.981	0.046±1.067	-0.78	473	0.44
	Ortalama	-0.026±0.960	0.044±1.036	-0.78	498	0.44
	EM	-0.038±0.990	0.050±1.053	-0.97	498	0.33
	Regresyon	-0.037±0.998	0.061±1.051	-1.06	498	0.29
%10 eksik	Pairwise	-0.060±0.995	0.050±1.071	-1.13	450	0.26
	Ortalama	-0.055±0.951	0.044±1.012	-1.13	498	0.26
	EM	-0.057±0.989	0.040±1.051	-1.07	498	0.29
	Regresyon	-0.060±0.993	0.054±1.054	-1.24	498	0.22
%15 eksik	Pairwise	-0.048±0.980	0.044±1.103	-0.82	429	0.41
	Ortalama	-0.043±0.913	0.028±1.020	-0.82	498	0.41
	EM	-0.045±0.974	0.044±1.060	-0.98	498	0.33
	Regresyon	-0.052±0.990	0.026±1.080	-0.85	498	0.40
%20 eksik	Pairwise	-0.085±1.010	0.080±1.074	-1.61	409	0.11
	Ortalama	-0.074±0.930	0.061±0.957	-1.60	498	0.11
	EM	-0.041±1.005	0.066±1.044	-1.17	498	0.24
	Regresyon	-0.031±1.022	0.088±1.044	-1.29	498	0.20

Çizelge 4.95'te 500 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X2 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.95- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X2 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.053±1.034</b>	<b>0.068±1.052</b>	<b>-1.30</b>	<b>498</b>	<b>0.20</b>
%5 eksik	Pairwise	-0.059±1.057	0.056±1.052	-1.19	473	0.24
	Ortalama	-0.055±1.017	0.054±1.039	-1.19	498	0.24
	EM	-0.058±1.037	0.064±1.052	-1.30	498	0.19
	Regresyon	-0.061±1.043	0.063±1.060	-1.32	498	0.19
%10 eksik	Pairwise	-0.032±1.029	0.060±1.041	-0.95	451	0.34
	Ortalama	-0.027±0.969	0.057±1.002	-0.95	498	0.34
	EM	-0.047±1.016	0.070±1.041	-1.27	498	0.20
	Regresyon	-0.049±1.014	0.066±1.039	-1.26	498	0.21
%15 eksik	Pairwise	-0.049±1.031	0.034±1.039	-0.83	427	0.41
	Ortalama	-0.044±0.958	0.028±0.959	-0.83	498	0.41
	EM	-0.055±1.026	0.050±1.025	-1.15	498	0.25
	Regresyon	-0.048±1.029	-0.073±1.059	-1.30	498	0.20
%20 eksik	Pairwise	-0.084±1.045	0.040±1.044	-1.19	405	0.23
	Ortalama	-0.072±0.940	0.028±0.944	-1.19	498	0.23
	EM	-0.054±1.025	0.080±1.054	-1.44	498	0.15
	Regresyon	-0.053±1.022	0.057±1.063	-1.18	498	0.24

Çizelge 4.96'da 500 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X3 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.96- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X3 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.037±1.011</b>	<b>0.041±1.043</b>	<b>-0.84</b>	<b>498</b>	<b>0.40</b>
%5 eksik	Pairwise	-0.053±1.003	0.054±1.040	-1.15	473	0.25
	Ortalama	-0.051±0.981	0.051±1.009	-1.15	498	0.25
	EM	-0.034±1.003	0.051±1.031	-0.94	498	0.35
	Regresyon	-0.044±1.000	0.065±1.029	-1.20	498	0.23
%10 eksik	Pairwise	-0.034±1.002	0.059±1.062	-0.96	451	0.34
	Ortalama	-0.030±0.954	0.054±1.010	-0.96	498	0.34
	EM	-0.041±1.001	0.042±1.047	-0.91	498	0.36
	Regresyon	-0.035±1.023	0.045±1.050	-0.87	498	0.39
%15 eksik	Pairwise	-0.017±1.031	0.019±1.061	-0.35	425	0.72
	Ortalama	-0.014±0.941	0.017±0.992	-0.35	498	0.72
	EM	-0.026±1.002	-0.037±1.044	-0.69	498	0.49
	Regresyon	-0.022±1.009	-0.033±1.056	-0.59	498	0.55
%20 eksik	Pairwise	-0.085±1.011	0.071±1.029	-1.54	408	0.13
	Ortalama	-0.073±0.925	0.054±0.922	-1.54	498	0.13
	EM	-0.063±1.009	0.036±1.040	-1.08	498	0.28
	Regresyon	-0.056±1.034	0.030±1.046	-0.92	498	0.36

Çizelge 4.97'de 500 birimlik tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde, X4 değişkenine ait farklı iki grubun (X5) karşılaştırılmasına ilişkin t testi sonuçları gösterilmektedir.

Çizelge 4.97- Tam veri seti ve farklı atama yöntemlerine göre tamamlanmış veri setlerinde X4 değişkenine ait farklı iki grubun karşılaştırma sonuçları

		Grup I ( $\bar{X} \pm S$ )	Grup II ( $\bar{X} \pm S$ )	t	sd	P
	<b>Tam Veri</b>	<b>-0.089±1.005</b>	<b>0.037±1.035</b>	<b>-1.38</b>	<b>498</b>	<b>0.17</b>
%5 eksik	Pairwise	-0.093±1.002	0.028±1.049	-1.28	473	0.20
	Ortalama	-0.089±0.972	0.025±1.027	-1.28	498	0.20
	EM	-0.090±1.001	0.035±1.035	-1.36	498	0.17
	Regresyon	-0.086±1.005	0.035±1.033	-1.33	498	0.18
%10 eksik	Pairwise	-0.112±1.000	0.039±1.017	-1.59	450	0.11
	Ortalama	-0.106±0.957	0.031±0.961	-1.59	498	0.11
	EM	-0.102±0.996	0.026±1.036	-1.41	498	0.16
	Regresyon	-0.099±1.005	0.014±1.041	-1.24	498	0.22
%15 eksik	Pairwise	-0.089±1.019	0.032±1.044	-1.22	428	0.23
	Ortalama	-0.080±0.932	0.024±0.982	-1.22	498	0.22
	EM	-0.083±0.993	0.025±1.019	-1.19	498	0.23
	Regresyon	-0.069±1.019	0.019±1.023	-0.96	498	0.34
%20 eksik	Pairwise	-0.095±0.977	0.052±1.044	-1.47	406	0.14
	Ortalama	-0.084±0.894	0.035±0.930	-1.47	498	0.14
	EM	-0.092±0.981	0.043±1.021	-1.51	498	0.13
	Regresyon	-0.098±0.983	0.074±1.061	-1.88	498	0.06

Yüksek korelasyonlu deęişkenlere sahip 500 birimlik veri setlerinin analizleri sonucunda %5'lik, %10'luk, %15'lik ve %20'lik eksik veri setlerinde farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise genel itibariyle EM atama yöntemi tam veri seti ile oldukça yakın sonuçlar vermiştir. Burada hemen hemen tüm atama yöntemlerinden birbirlerine ve tam veri setine yakın sonuçlar elde edilmiştir.



## 5- TARTIŞMA VE SONUÇ

Türetilen tüm veri setlerinin analizleri sonucunda, kullanılan atama yöntemlerinin (Listwise, Pairwise, Ortalama, EM ve Regresyon) mevcut teorik bilgilerine paralel olarak sonuçlar elde edilmiştir. Atama yöntemleri aralarındaki farklar, yöntemlere göre birbirlerinden önemli düzeyde farklılaşmıştır. Örneğin tüm düşük ve yüksek korelasyonlu değişkenlere sahip eksik veri setlerinde, Ortalama ve Pairwise atama yöntemi ile elde edilmiş veri setlerinin ortalamaları eşit iken, bu yöntemlere ilişkin veri setlerinin standart sapmaları (S) düşük çıkmıştır. Pairwise ve Ortalama atama yöntemlerine ait elde edilen t testi sonuçları ise birbirine oldukça yakın bulunmuştur. Genel olarak, Ortalama atama yöntemi ile elde edilen istatistikler, Pairwise yöntemi ile elde edilen istatistiklere göre az da olsa küçük ve buna bağlı olarak, olasılık değeri Pairwise'a göre yüksek bulunmuştur ( $P < 0.05$ ). Ayrıca Ortalama atama yöntemi ile tamamlanan veri setlerine ait regresyon analizlerinde, belirleyicilik katsayıları tam veri setlerine göre oldukça düşük çıkmış, regresyonun açıklayıcılığı oldukça düşmüştür.

Ortalama atama yönteminde, değişkene ait ortalama ve medyan değerleri sabit kalmakta ve standart sapmalar küçülmektedir. Bu durumda, veri setlerindeki değişkenlerin dağılım yapıları değişmekte ve grup ortalamaları arasında gerçekte olan farklar standart sapmanın küçülmesiyle bir anda önemsiz olarak görülebilmektedir. Bu nedenle bu atama yöntemlerinin kullanılmasında dikkatli davranılması yerinde olacaktır. Little ve Rubin (22, 24) yaptıkları çalışmalarda benzer sonuçlara ulaşmışlardır.

50 birimlik veri setlerinin analizlerinde ortalamaların tahmininde genel itibariyle EM atama yöntemi, diğer atama yöntemlerine göre daha iyi sonuçlar vermiş olmasına rağmen, düşük örnek hacimlerine sahip veri setlerinde bu durum yeterli bir sonuç olarak belirlenememiştir. Bunun nedeni veri setindeki birim sayısının yetersizliğidir. Özellikle  $n=50$  veri seti için %15 ve %20 eksik gözlemlili örneklerde eksik veri analizleri için tam bir yorum yapmak bu durumda oldukça güç olmaktadır. Regresyon analizleri sonucunda



elde edilen genel tabloda da sonuçlarda tutarsızlıklar saptanmıştır. Düşük korelasyonlara sahip veri setinde, Listwise yöntemi ile tamamlanan veri seti diğer yöntemlere göre biraz daha tam veri setine yakın sonuçlar vermesine rağmen, eksik veri yüzdesi arttığında değişkenin toplumdaki dağılımı önemli düzeyde tepeleşme göstermektedir. Bu durumun gerçek yapıya uyumluluğu tartışmalıdır. Gerçekte grup ortalamaları arasında varolmayan istatistiksel farklılıklar, Regresyon ve Ortalama atama yöntemlerinde ortaya çıkmaktadır. Bu durum ise araştırmacının hatalı kararlar vermesine sebebiyet verebilir. Little ve Rubin (22), Schafer (36, 37) az birim içeren ve eksik veri oranı yükseldikçe eksik veri analizi veri setlerinde hatalı çıkarsamalara yol açtığını belirtmişlerdir.

100 birimlik veri setlerinin analizlerinde de grup ortalamalarının benzerliklerinde 50 birimlik verilerde olduğu gibi EM atama yöntemi, diğer atama yöntemlerine göre tam veri setine daha yakın sonuçlar vermiştir. Fakat 50 birimlik veri setlerindeki küçük örnek hacmi problemi burada da geçerlidir. Özellikle veri eksikliğinin %15 ve %20 olduğu durumlarda sonuçlar belirsiz bir durum almaktadır. Sonuçlar tam veri seti ile karşılaştırıldığında tutarsız bir hale gelmektedir. Regresyon atama yöntemi ile tamamlanmış veri setlerinde yapılan regresyon analizlerinde belirleyicilik katsayısı tam veri setinde olduğundan daha yüksek çıkma eğilimindedir. Bu durum ise değişkenler arasındaki korelasyon katsayılarına bağlı olarak artmaktadır. Çünkü, Regresyon atama yönteminde değişkenler arasındaki korelasyonlar artmaktadır. Gruplar arasındaki farklılıkların karşılaştırılmasında genel olarak EM atama yöntemi tam veri setine yakın değerler vermiştir.

200 birimlik veri setlerinin analizlerinde 50 ve 100 birimlik veri setlerine göre daha tutarlı sonuçlar elde edilmiştir. Ortalamaların tahmininde EM ve Pairwise atama yöntemleri eniyi sonuçları vermiştir. Özellikle EM atama yöntemi, gerek regresyon analizlerinde gerekse t testi sonuçlarında daha güvenilir bir yöntem görünümündedir. Yüksek korelasyonlara sahip veri setlerinin analizinde EM atama yöntemi diğer yöntemlere göre daha güvenilir bir tablo çizmiştir. Özellikle %5 ve %10 eksik veri setlerinde, tam veri setlerine daha yakın sonuçlar elde edilmiştir. %15'lik eksik veri setinin atama yöntemleri ile tamamlanması sonucunda, farklı iki gruba ait ortalamalar

arasındaki farkların saptanmasında genel itibariyle EM atama yöntemi benzer sonuçlar vermesine rağmen, genel olarak tüm yöntemlerden elde edilen anlamlılık düzeyleri (p) artmış ve t değerleri düşmüştür.

300 birimlik veri setlerinin analizlerinde, düşük korelayonlu veri setinde, X4 değişkende tam veri setinde gruplar arasında fark bulunmazken Regresyon atama yönteminde fark bulunması oldukça düşündürücüdür. Fakat veri setlerindeki birim sayıları arttıkça, atama yöntemleri ile elde edilen veri setleri ile tam veri setleri arasındaki uçurumlar gittikçe kapanmaktadır. 50, 100 ve 200 birimlik veri setlerine göre, daha tutarlı sonuçlar elde edilmiştir. Farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise EM atama yöntemi tam veri setinden elde edilen değerlere oldukça yakın sonuçlar vermiştir. Özellikle EM atama yöntemi, gerek regresyon analizlerinde, gerekse t testi sonuçlarında daha güvenilir bir yöntem görünümü almaktadır. Yüksek korelasyonlara sahip veri setlerinin analizinde EM atama yöntemi diğer yöntemlere göre daha güvenilir bir tablo çizmiştir. Özellikle %5 ve %10 eksik veri setlerinde tam veri setlerine daha yakın sonuçlar elde edilmiştir. %15'lik eksik veri setinin atama yöntemleri ile tamamlanması sonucunda, farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında, genel itibariyle EM atama yöntemi benzer sonuçlar vermesine rağmen, genel olarak tüm yöntemlerden elde edilen anlamlılık düzeyleri (p) artmış ve t değerleri düşmüştür.

400 birimlik düşük ve yüksek korelasyonlu değişkenlere sahip veri setlerinin analizlerinde; 50, 100, 200 ve 300 birimlik veri setlerine göre daha anlamlı ve tutarlı sonuçlar elde edilmiştir. Farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise EM atama yöntemi, tam veri setinden elde edilen sonuçlara oldukça yakın değerler vermiştir. Özellikle EM atama yöntemi, gerek regresyon analizlerinde, gerekse t testi sonuçlarında daha güvenilir bir yöntem görünümü almıştır. Her ne kadar Regresyon atama yönteminin belirleyicilik katsayıları tam veri setine benzerlik gösterse de regresyon katsayılarında ve bu katsayıların anlamlılık seviyelerinde, gerçekten oldukça uzak tahminler yapmaktadır. Yüksek korelasyonlara sahip veri setlerinin analizinde, EM atama yöntemi diğer yöntemlere göre daha güvenilir bir görünümde dir.

500 birimlik veri setlerinin analizlerinde; farklı örnek hacimlerine sahip veri setlerine göre, daha anlamlı ve tutarlı sonuçlar elde edilmiştir. Farklı iki gruba ait ortalamalar arasındaki farkların saptanmasında ise EM atama yöntemi, tam veri setinden elde edilen sonuçlara oldukça yakın değerler vermiştir. Özellikle EM atama yöntemi, gerek regresyon analizlerinde gerekse t testi sonuçlarında daha güvenilir bir yöntem görünümü çizmiştir. Örnekteki birim sayılarının 500 birime yaklaşması hem tahminleri tutarlı bir hale getirmiş, hem de farklı atama yöntemleri ile elde edilen tahminleri birbirine yaklaştırmıştır. Özellikle EM atama yöntemi, düşük ve yüksek hacimli örneklerde diğer atama yöntemlerine göre daha başarılı bir tablo çizmiştir. 300 birim ve üzeri ve değişkenlerindeki eksik veri sayıları, verilerin tamamının %15'inden az olan veri setlerinde ( $n_{\text{mis}} < 0,15$ ) yapılacak olan tahminlerde, EM yönteminin kullanılması daha güvenilir tahminlerin elde edilmesini sağlayacaktır. Her ne kadar Regresyon atama yönteminin belirleyicilik katsayıları tam veri setine benzerlik gösterse de, regresyon katsayılarında ve bu katsayıların anlamlılık seviyelerinde gerçekten uzak tahminler elde edilmektedir. Regresyon katsayılarının anlamlılık düzeyleri dikkate alındığında, EM atama yöntemi daha duyarlı sonuçlar vermektedir. Özellikle yüksek korelasyonlu değişkenlere sahip veri setlerinde, EM atama yöntemi diğer yöntemlere göre daha güvenilir sonuçlar ortaya koymuştur.

Regresyon atama yönteminde, atama esnasında değişkenler arasındaki korelasyonlar büyümekte ve buna bağlı olarakta gerçekte anlamsız olan bazı değişkenlere ait regresyon katsayıları, anlamlı bir görünüm sergilemektedir. Aynı zamanda belirleyicilik katsayıları ( $R^2$ ) olması gerekenden yüksek çıkararak hatalara yol açabilmektedir. Bunlar, Regresyon atama yönteminin en önemli sakıncalarındandır. Gerçekte olmamasına rağmen, farklı gruplara ait ortalamalar arasındaki farklar da anlamlı çıkabilmektedir.

Yıllarca araştırmacılar kullandıkları paket programlarda eksik veri problemine, Pairwise yöntemini kullanarak, veri setine ilişkin korelasyon ve kovaryans matrisini hesaplayarak yaklaşmışlar ve de bu matrisi istatistiksel analizler için girdi matrisi olarak kullanmışlardır. Fakat analizlerde kullanılan bu tür matrislerin özdeğerleri 0'dan küçük olabilmekte ve bazı korelasyonlar, birimlerin farklı alt kümelerinden hesaplanmış

olabilmektedir. Bazı arařtırmacılar ise EM veya Regresyon yöntemlerini tahmin istatistikleri veya da veri atama yöntemi için kullanmışlardır. Yapılan benzetim çalışmaları, Pairwise tahminlerinin, EM yönteminden elde edilmiş tahminlere göre daha kötü sonuçlar verdiğini göstermiştir (13).

Cool (2000), yaptığı çalışmada eksik veri ile başa çıkmak için 4 seçim önermektedir. Fakat, bu yöntemlerden, Listwise ve Pairwise silme yöntemlerinin, örnek hacmini düşürdüğünü, bu durumun da toplum parametre tahminlerinin kesinliğinin düşmesine neden olduğunu belirtmektedir. Örnek hacmindeki düşüşün, aynı zamanda istatistiksel önemlilik testlerinin gücünü düşürdüğünü ve bu durumun karar vermede tutarsız sonuçlara neden olduğunu açıklamaktadır. Çoklu bağımlılığın (multicollinearity) düşük olduğu durumlarda ise, Ortalama atama ve Regresyon atama yöntemlerinin eşit derecede etkili olduğu, Raymond ve Roberts (1987) tarafından belirtilmiştir (7). Ayrıca Cool (2000), Ortalama atama yönteminin en önemli avantajını, örnek hacmini sabit tutması ve buna bağlı olarak istatistiksel gücün değişmemesi olarak açıklamıştır (7).

Sağlık alanında literatürde yapılmış çalışmalar incelendiğinde, benzer yönde sonuçların bulunduğu görülmektedir. Graham ve arkadaşları (1994) yaptıkları uyuşturucu ile ilgili bir çalışmada, EM atama yöntemin diğer atama yöntemlerine göre daha iyi sonuçlar verdiğini belirtmişlerdir (10).

Sağlık alanındaki bir başka çalışma ise Kneipp ve McIntosh (2001) tarafından yapılmıştır. Kneipp ve McIntosh, eksik veri problemini hemşirelik arařtırmalarında ele almışlardır. Bu arařtırmada veri setine çoklu atama yöntemini uygulamışlar ve elde ettikleri tamamlanmış veri seti ile istatistiksel analizleri yapmışlardır. Bu çalışmada Listwise, Pairwise ve tekli (Ortalama, Medyan vb) atama yöntemlerinin veri yapılarını korumadıklarını en iyi atama yönteminin MI (çoklu atama) olduğunu savunmuşlardır (19).

Laurikala ve arkadaşları (2001), kadın ürünler sisteminin tanı, teşhis ve tedavisi için elde ettikleri veri setindeki eksik veriler için, Ortalama, Regresyon ve EM atama yöntemlerini kullanmışlardır. Eksik verilerin yerine tahmin yöntemleri ile atama

yaptıktan sonra veri setine istatistiksel analizleri uygulamışlardır. Bu çalışmada Ortalama yöntemi, Regresyon ve EM yöntemine göre daha iyi sonuçlar vermiştir (20). Yapılan analizler sonucunda bu durumun, Laurikala ve arkadaşlarının yaptıkları araştırmaya has bir durum olduğu sonucuna ulaşılabilir.

Laurikkala ve arkadaşları (1999) yaptıkları bir diğer çalışmada, otolojik bir veri setinde bulunan eksik verilerin tahmini için Ortalama, EM ve Regresyon atama yöntemlerini kullanmışlardır. Bu üç atama yöntemine ilişkin elde edilen sonuçlar birbirlerine benzerlik göstermiştir (21).

West ve Dawson (2002), tekrarlı ölçümlere dayalı uzunlamasına (longitudinal) çalışmalarında, eksik verilerin tahminlenmesinde EM atama yöntemini kullanmışlardır. Bu çalışmada; öncelikle elde edilen veri seti (available cases) sonrasında ise EM yöntemi ile eksik verileri tamamlanan (complete cases) veri seti analiz edilmiş ve bu sonuçlar karşılaştırılmıştır. En uygun sonuçları, EM atama yöntemi ile elde edilen veri seti vermiştir (47).

Crawford ve arkadaşları (1995), yaptıkları çalışmada, saha araştırmalarında karşılaşılan eksik verilere sahip sonuç değişkenlerinin atama yöntemleri ile tahminini araştırmışlardır. Bu yöntemlerden, Ortalama atama yönteminin, atama yöntemi olarak uygun sonuçlar vermediği, fakat modele dayalı atama yöntemlerinin atama için uygun olduğu fikrini savunmuşlardır (8).

Musil ve arkadaşları (2002), 492 birimlik eksik verilere sahip bir veri setinde Ortalama, Listwise, Regresyon ve EM atama yöntemlerini tahmin yöntemi olarak kullanmışlardır (96 gözlem değerine ait eksik verileri tahmin etmeye çalışmışlardır.). Tüm atama yöntemlerin sınırlılıklarından bahsetmişlerdir. Ortalama, Regresyon ve EM atama yöntemlerinin birbirlerine benzer sonuçlar verdiğini ve en iyi atama sonuçlarının ise Listwise yöntemi ile elde edildiğini belirtmişlerdir (25).

Azen ve arkadaşları (1998), bir benzetim çalışması yapmışlardır. Bu çalışma, elde ettiğimiz sonuçlarla paralellik göstermektedir. Burada; tam veri seti, tüm eldeki veriler (Pairwise) ve EM atama yöntemlerinin performansları karşılaştırılmıştır. Ayrıca, veri setindeki değişkenlerin normal, karışık normal ve lognormal olduğu durumlar için

analizler yapılmıştır. %5'lik eksik veri yapısına sahip veri setlerinde, EM ve tam verilerin analiz sonuçları birbirlerine benzer sonuçlar vermişlerdir. Eksiklik %25'lere çıktığında ise EM algoritması genelde en iyi tahminleri vermiştir. Fakat araştırmacılar, Pairwise yöntemi ile yapılan analizlerin, bu çalışmada her ne kadar iyi sonuçlar vermese de, hem daha güvenli hem de mantıklı olduğunu savunmuşlardır. Sonuç olarak, eksik verilerin tahmininde optimal çözümün EM algoritması ile olduğu sonucunu elde etmişlerdir (4).

Cool (2000), yaptığı çalışmada, Listwise, Pairwise, Ortalama ve Regresyon atama yöntemlerini karşılaştırmıştır. SPSS paket programını kullanarak yaptığı analizlerde Listwise ve Pairwise yöntemlerinin örnek hacimlerini düşürdüğünü ve dolayısıyla istatistiksel testlerin güçlerinin de düştüğünü açıklamıştır. Ortalama ve Regresyon atama yöntemlerinde ise korelasyonların arttığını açıklamıştır. Ortalama atama yönteminin en önemli avantajının, örnek hacminin korunması olduğunu ve buna bağlı olarak da istatistiksel gücün korunduğunu belirtmiştir (7).

Pigott (2001), yaptığı çalışmada eksik veri analizlerini ele almıştır. Bir çok araştırmacının, atama yöntemi olarak ad-hoc (Listwise, Pairwise veya Ortalama atama yöntemleri) yöntemlerini kullandıklarını belirtmiştir. Bu yöntemlerin, basit yöntemler olduğunu modele dayalı yöntemlerin, örneğin EM atama yönteminin eksik verilerin tahmin edilmesinde bir takım zorlukları olmasına rağmen, iyi sonuçlar verdiğini açıklamıştır. Uygulama olarak, sağlık alanında astımlı öğrenciler üzerinde bir çalışma yapmıştır (33).

Oğuzlar (2001), dünya bankasından elde ettiği veri setine atama yöntemlerini uygulamıştır. Araştırdığı veri setinde bulunan değişkenlerinde, çok fazla eksik veri bulunduğundan, %60 ve üzeri eksik veri içeren değişkenleri araştırma kapsamı dışında bırakmıştır. Listwise, Pairwise, Regresyon ve EM atama yöntemlerini kullanmıştır. EM ve Regresyon atama yöntemi ile elde edilen veri setlerini istatistiksel analizlerde kullanmıştır (27).

Zhou ve arkadaşları (2001), çalışmalarında, halk sağlığı araştırmalarında sıklıkla karşılaşılan eksik veri problemini ele almışlardır. Bu araştırmalarda sıklıkla kullanılan,

Ortalama ve Medyan atama yöntemlerinin oldukça basit yöntemler olduğunu açıklamışlardır. Çalışmalarında, iki gerçek veri seti ve bir de gerçek çalışmaya dayalı benzetim ile tüm atama yöntemlerini karşılaştırmışlardır. Araştırma sonucunda, çoklu atama yöntemlerinin, tekli atama yöntemlerine daha iyi tahminler verdiğini belirtmişlerdir (49).

Patrician (2002), yaptığı çalışmada, genelde araştırmalarda geleneksel yaklaşımlar olan Listwise ve Ortalama atama yöntemlerinin eksik veri analizlerinde yeterli olmadığını belirtmiştir. Bununla birlikte son zamanlarda geliştirilen yeni yöntemlerin parametre tahminlerini, standart hataları ve test istatistiklerini geliştirdiklerini belirtmişlerdir (32).

Streiner (2002), yaptığı çalışmada; elinde bulunan veri setini; Ortalama, Regresyon, Regresyon+hata ve çoklu atama yöntemlerini kullanarak tamamlamıştır. Atama yöntemlerinden elde ettiği sonuçları dikkate alarak, en iyi atama yöntemini çoklu atama yöntemi olarak belirtmiştir. Orjinal verilere yakınlığına göre en iyi tahmini Ortalama ve Regresyon yöntemleri vermiştir. Regresyon+hata yönteminden elde edilen sonuçlar gerçek değerlerden uzak bulunmuştur (42).

Sonuç olarak, sadece sağlık alanında bir çok araştırmada değil iktisat, sosyoloji, psikoloji vb. diğer pek çok bilim alanda da yapılan çalışmalarda eksik veri durumuyla karşılaşmaktadır. Eksik veriyi çözümlenmenin en önemli ve de kolay yolu veri toplama aşamasında bu durumu çözmektir. Eğer eksik veri durumu kaçınılmaz ise, eksik veri için kestirimde bulunabilecek yararlı ortak değişkenlerin toplanmasında fayda vardır. Eksik verinin önemsenmeyebilir (TROC ve/veya ROK) ve önemsenebilir sınıflamalarından hangisine girdiğinin ayırımını iyi yapmak, doğru bir modelleme ve çözümlenme yapılabilmesi bakımından oldukça önemlidir. Bu araştırmada yapılan tüm benzetimsel analizler ışığında literatürle benzer sonuçlar elde edilmiştir. Listwise ve Pairwise atama yöntemlerinde birim sayıları düşmüş, elde edilen analiz sonuçları gerçek istatistiklerden uzaklaşmıştır. Özellikle küçük örneklerde tutarsız sonuçlar elde edilmiştir. Ortalama atama yönteminde ise örnek hacmi sabit kalmış, fakat değişkenlere ilişkin varyans değerleri küçülmüştür. Bu ise değişkene ilişkin dağılımı bozulmasına neden olmuştur. Örnek hacminin şişirilmesi ile de testin gücü aslında yüksek gibi

görünmesine rağmen, gerçekte bu durum hataya sebebiyet verebilmektedir. Gerçekte grup ortalamaları arasında var olabilecek farklar, bu fark gerçekte bulunmasına rağmen varyansların küçülmesiyle önemliliklerini yitirebilmektedir. Regresyon atama yönteminde ise değişkenler arasındaki korelasyonlar, gerçekte var olan korelasyonlara göre yüksek bulunmuş, bu ise değişkenler arasında var olan ilişkilerin olduğundan daha anlamlı ve önemli çıkmasına sebebiyet vermiştir. Özellikle, küçük örneklerde bu atama yönteminde tutarsız sonuçlar gözlenmiştir. EM atama yöntemi ise diğer atama yöntemlerine göre en iyi sonuçları vermiştir. EM atama yöntemi ile elde edilmiş veri setlerine ait istatistikler, gerek düşük korelasyonlu gerekse yüksek korelasyonlu değişkenlere sahip veri setlerinde, orijinal (tam, complete) veri setleri ile benzer sonuçlar vermiştir. Veri setlerindeki eksik olma durumlarında, özellikle %10 ve daha az eksik olma durumunda EM atama yöntemi tam verilere oldukça yakın sonuçlar vermiş, %10 ve daha fazla eksik veri olduğuna ise diğer atama yöntemlerine üstünlük sağlamıştır. Özellikle, örnek hacmi yüksek ( $n > 200$ ) veri setlerinde sonuçlar EM atama yönteminde tam verilerle oldukça benzer sonuçlar vermiştir.

Araştırmalarda eksik veri problemi ile sık sık karşılaşılmaktadır. Veri setlerindeki eksik veri probleminin çözümü için, değişkenlere ait elde edilen ölçümlerin, dağılım formlarının, örnek hacminin, veri setindeki eksik verilerin yüzdesinin önemsenebilir olup olmadığının, bu eksikliğin nereden kaynaklandığının belirlenmesi gibi sorulara cevap aranması gerekmektedir. Veri setindeki eksik veriler yerine, gerçekteki değerlerine benzer değerlerin atanması işlemi aslında oldukça zor ve ayrıntılı istatistiksel işlemler gerektirir. Paket programların ilgili bölümlerinden yararlanılarak değer atama işlemi aslında çok basit bir işlem gibi görünmesine rağmen, atama işleminin ardındaki analizler ve varsayımlar oldukça karmaşık ve kompleks bir yapıya sahiptir.

Yaptığı araştırmada eksik veri problemi ile karşılaşan bir araştırmacı, verilerin taşıdığı mekanizmaların belirlenmesi ve uygun atama yönteminin seçilmesi için mutlaka bir istatistikçi ya da biyoistatistikçiye başvurmalı ve bu konu ile ilgili danışmanlık hizmeti almalıdır. Bilgisi ve deneyimi ne kadar geniş olursa olsun istatistik eğitimi almamış bir kişinin her konuda uzman olması 'bilimin doğasına' aykırı bir durumdur.



Bu nedenle, bir araştırma tek kişilik bir işlem olamaz. Multidisipliner bir ekip çalışmasını gerektirir. Her araştırma, konu ile değişik branşlardan oluşan ekip tarafından planlanmalı, yürütülmeli ve sonlandırılmalıdır (28, 43, 44).

Aslında doğru olan; araştırmanın planlama aşamasından itibaren, araştırma sonlanana kadar bir istatistikçi ya da biyoistatistikçinin araştırma protokolü içerisinde yer alması ve oluşacak bu gibi olumsuz durumları önceden görerek, tedbirinin alınmasını sağlamasıdır. Araştırmanın en önemli aşamaları, veri toplama ve bu verilerin istatistiksel analizi aşamalarıdır. Bu aşamaları kontrol altına alınan araştırmalar, daha geçerli ve güvenilir, tutarlı, temeli güçlü bilimsel bilgiler üretecek ve yasalar ortaya koyacaktır.



## 6. BİR UYGULAMA

Türetilmiş verilerde yapılan araştırma sonuçları dikkate alınarak, EM atama yöntemi sağlık alanından elde edilen gerçek bir veri setine uygulanmıştır. Bu çalışmada uygulama verileri; 2001 ile 2003 yılları arasında, Osmangazi Üniversitesi Tıp Fakültesi Eğitim ve Uygulama Hastanesi Kadın Doğum ve Genel Cerrahi Polikliniklerine meme kanseri açısından tarama amacıyla başvuran ve fizik muayenesi negatif, asemptomatik 214 hastanın, Radyoloji Anabilim Dalı'ndaki, Mammografi ünitesindeki hasta tanıtım formları ve ek bilgi formlarındaki bilgileri olarak alınmıştır. Veriler kişi kimliği belirtilmeksizin ve Radyoloji Anabilim Dalı Başkanının izni ile kullanılmıştır.

Bu veri setinde bulunan değişkenler;

- YAŞ : Hastanın kliniğe başvuru anındaki yaşı  
AD\_YAS : Hastanın ilk adet yaşı  
DOG\_SAY : Hastanın yaptığı canlı doğum sayısı  
ID\_YAS : Hastanın ilk doğum yaşı  
ULTRA\_S : Ultrason sonucu (0=Normal; 1=Benin bulgular)  
SIGARA : Hastanın düzenli olarak sigara kullanıp kullanmadığı  
(0=Kullanmıyor; 1=Kullanıyor)  
TES : Toplam emzirme süresi  
(0=Emzirmemiş; 1=6 aydan az emzirmiş; 2=6 ay ile 2 yıl  
arasında emzirmiş; 3=2 yıl ve daha fazla emzirmiş)

Bu veri setinde meme kanserine neden olduğu düşünülen, hastanın; yaşı, ilk adet yaşı, yaptığı doğum sayısı, ilk doğum yaşı, sigara kullanımı ve emzirme süresi gibi değişkenler analiz kapsamına dahil edilmiştir. Veri setinde bulunan yaş, doğum sayısı, ultrason sonucu, sigara kullanımı ve toplam emzirme süresi değişkenlerine ilişkin geçerli ve eksik olma durumları çizelge 6.1'de görülmektedir. Bu değişkenlere ilişkin incelenen veri setinde, hiç eksik veri bulunmamaktadır.

214 kadından elde edilmiş değişkenlere ait alt kategoriler, her bir alt kategoriye düşen birim sayıları ve bunlara ait yüzde değerleri ise çizelge 6.2 de görülmektedir.

Çizelge 6.1: Değişkenlere ilişkin geçerli ve eksik gözlem sayısı

	<b>DOG_SAY</b>	<b>ULTRA_S</b>	<b>SIGARA</b>	<b>TES</b>	<b>YAŞ</b>
Geçerli	214	214	214	214	214
Eksik	0	0	0	0	0

Çizelge 6.2: Toplam 214 kadına ait değişkenler, alt kategorileri, sayı ve yüzdeleri

<b>Değişkenler</b>	<b>Kategoriler</b>	<b>Sayı</b>	<b>Yüzde</b>
<b>Doğum Sayısı</b>	0	15	7.0
	1	18	8.4
	2	125	58.4
	3	40	18.7
	4	16	7.5
	<b>Toplam</b>	<b>214</b>	<b>100.0</b>
<b>Ultrason Sonucu</b>	Negatif	111	51.9
	Pozitif (Benin Bulgular)	103	48.1
	<b>Toplam</b>	<b>214</b>	<b>100.0</b>
<b>Sigara Kullanımı</b>	Kullanmıyor	166	77.6
	Düzenli Kullanıyor	48	22.4
	<b>Toplam</b>	<b>214</b>	<b>100.0</b>
<b>Toplam Emzirme Süresi</b>	Emzirmemiş	16	7.5
	< 6 ay	32	15.0
	6 ay-2 yıl	86	40.2
	> 2 yıl	80	37.4
	<b>Toplam</b>	<b>214</b>	<b>100.0</b>

Fakat, 214 hastaya ait verilerden bazılarında, ya herhangi bir sebepten ulaşılamamış, ya da kayıt altına alınırken atlanmıştır. Bu değişkenlere ait eksik olma sayıları ve bunlara ait yüzdeler çizelge 6.3'te verilmiştir.

Çizelge 6.3- Hastalardan elde edilen değişkenlerin eksik olma sayı ve yüzde değerleri

Değişken	n	Eksik		Aşırı Değerler	
		Sayı	Yüzde	Düşük	Yüksek
AD_YAS	185	29	13.6	1	5
ID_YAS	174	40	18.7	0	6

Değişkenlerdeki eksik olma yapısı ise çizelge 6.4'te gösterilmiştir. YAS, DOG\_SAY, SIGARA ve TES değişkenlerinde eksik gözlem değeri bulunmamaktadır. Elde edilen sonuçlara göre AD\_YAS değişkeninde toplam 29, ID\_YAS değişkeninde ise toplam 40 kadına ait gözlem değerine ulaşılammıştır. Bu iki değişkende de aynı anda eksik değeri olan 9 kadın (birim) bulunmaktadır. Tüm değişkenlere ait gözlem değerleri tam olan birim sayısı ise 154'tür. Yapılan Little'nin TROK testi sonuçlarına göre veri setinin TROK yapıda olduğu anlaşılmıştır (Little's MCAR test  $\chi^2=4.019$ ;  $P>0.05^{ns}$ ).

Çizelge 6.4: Veri setine ilişkin eksik olma paterni

Değişkenlerin eksik Birim Sayısı	YAS	DOG_SAY	SIGARA	TES	AD_YAS	ID_YAS
154						
20					X	
9					X	X
31						X

Listwise, pairwise ve EM yöntemlerine göre değişkenlere ilişkin ortalama, standart sapma değerleri ve bu değerlerin hesaplandığı birim sayıları çizelge 6.5'te verilmiştir.

Çizelge 6.5: Değişkenlere ilişkin ortalama ve standart sapma değerleri

	YAS $\bar{X} \pm s(n)$	AD_YAS $\bar{X} \pm s(n)$	ID_YAS $\bar{X} \pm s(n)$
Listwise	46.90±3.224 (154)	13.32±1.431 (154)	22.03±3.329 (154)
Pairwise	46.74±3.199 (214)	13.28±1.353 (185)	22.14±3.432 (174)
EM	46.74±3.199 (214)	13.28±1.353 (214)	22.15±3.431 (214)

Eksik ve EM atama yöntemi ile tamamlanan veri setlerinde yapılan t test istatistikleri, karşılaştırmalı olarak çizelge 6.6'da gösterilmektedir. Çizelge 6.6 incelendiğinde, YAS değişkeni, her iki veri setinde de birbirine eşit olduğu için (eksik veri içermediği için) tüm sonuç istatistikleri birbirine eşit elde edilmiştir. Adet yaşı (AD\_YAS) değişkeninde, negatif ve pozitif gruplarda az da olsa değişim gözlenmiştir. EM yöntemi ile elde edilen veri setinin analizi sonucunda, t değerlerinde 0.01'lik bir artış gözlenmiştir. Asıl değişim, t testi hesaplamalarında kullanılan serbestlik derecesinde gözlenmektedir. Eksik veri setinde 183 olan serbestlik derecesi, EM atama yönteminde 212 olarak elde edilmiştir. Bu durum ise, t değeri sabit olmasına rağmen örnek hacimlerini yükselttiğinden testin gücünü arttırmıştır. Daha fazla örnekle (sample size) bu değerler elde edildiği için tutarlılık artmış, yapılan testin topluma genellenebilirliği az da olsa artmıştır. Aynı durum ilk doğum yaşı (ID\_YAS) değişkeni içinde geçerlidir (Araştırma kapsamında 199 kadın, en az bir kez canlı doğum yapmıştır). Bu değişkene ilişkin yapılan t testlerinde de, birim sayılarının artmasına bağlı olarak, serbestlik dereceleri artmıştır ( $sd_E=172$  ;  $sd_{EM}=197$ ).

Çizelge 6.6: Eksik ve EM atama yöntemi ile tamamlanan veri setlerinde yapılan t testleri

		Tümör - ( $\bar{X} \pm S$ )	Tümör + ( $\bar{X} \pm S$ )	t	sd	P
Eksik	YAS	46.70±3.329	46.78±3.068	-0.17	212	0.866
	AD_YAS	13.47±0.921	13.07±1.675	2.06	183	0.041*
	ID_YAS	21.17±2.877	23.26±3.687	-4.19	172	0.000***
EM	YAS	46.70±3.329	46.78±3.068	-0.17	212	0.866
	AD_YAS	13.45±0.855	13.09±1.566	2.07	212	0.040*
	ID_YAS	21.30±2.714	23.13±3.455	-4.19	197	0.000***

Bir çok çalışmada hastaların yaşlarının meme kanserinde önemli risk faktörü olduğu saptanmıştır (34). Fakat bu çalışmada yaş değişkeni bir risk faktörü görünümünde bulunmamıştır ( $t=2.07$ ;  $sd=212$ ;  $P>0.05$ ). Ayrıca, erken yaştaki doğumlar



## KAYNAKLAR DİZİNİ

1. Akgül A.: Tıbbi Araştırmalarda İstatistiksel Analiz Teknikleri, Yükseköğretim Kurulu Matbaası, Ankara, 1997.
2. Alpar R.: Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1, Nobel Yayın Dağıtım, Ankara, 2003.
3. Alzola C., Harrell F.: An Introduction to S and The Hmisc and Design Libraries, University of Virginia School of Medicine, Charlottesville Va, USA, 2003.
4. Azen S. P., Van Guilder M., Hill M. A.: Estimation of Parameters and Missing Values Under a Regression Model with Non-normally Distributed and Non-randomly Incomplete Data, Stat. Med., 8 (2), 217-28, 1998.
5. Bal C.: Tedavi Sonrası İzlem Verilerinin Cox Regresyon Aracılığı ile İncelenmesi, Yüksek Lisans Tezi, OGÜ. Sağlık Bilimleri Enstitüsü, 1997.
6. Bekiroğlu N: Uzunlamasına Çalışmalarda Kayıp Verilerin Çözümleme Yöntemleri ve Bir Uygulama, 5. Ulusal Biyoistatistik Kongresi Bildiri Kitabı, Eskişehir, 2000.
7. Cool A. L.: A Review of Methods for Dealing with Missing Data, Annual Meeting of the Southwest Educational Research Association, Dallas, 2000.
8. Crawford S. L., Tennstedt S. L., McKinlay J. B.: A Comparison of Analytic Methods for Non-random Missingness of Outcome Data, J. Clin. Epidemiology, 48 (2), 209-19, 1995.
9. Diggle P. J., Liang K., Zeger S. L.: Anaysis of Longitudinal Data, Oxford University Press. Inc, 2000.
10. Graham J. W., Hofer S. M., Piccinin A. M.: Analysis with Missing Data in Drug Prevention research, Nida research Monograph, 142, 13-63, 1994.
11. Hagenaars J. A.: Categorical Longitudinal Data, Sage Publications Inc., 1990.

## KAYNAKLAR DİZİNİ (Devam)

12. Hesterberg T. C.: A Graphical Representation of Little's Test for MCAR. Technical Report No. 94, Research Department, Insightful Corporation, 1700 Westlake Ave. N., Suite 500, Seattle, WA 98109. <http://www.statsci.com/Hesterberg/little7.ps.gz> 1999. (Download 18.04.2003)
13. Hill M.: SPSS Missing Value Analysis 7.5, SPSS Inc., Chicago, 1987.
14. Horton N. J., Lipsitz S. R.: Multiple Imputation in Practise: Comparison of Software Packages for Regression Models with Missing Variables, *The American Statistician*, 55, 3, 244-252, 2001.
15. Howell D. C.: Treatment of Missing Data, (Last revised: December 23, 2002) [http://www.uvm.edu/~dhowell/Statpages/More\\_stuff/Missing\\_data/Missing.html](http://www.uvm.edu/~dhowell/Statpages/More_stuff/Missing_data/Missing.html). (Download 13.04.2003)
16. Hox J., Leeuw E.: Ad Hoc Solutions for missing Data Lecture notes, Köln, March, 1998.
17. Hunt L., Jorgenson M.: Mixture Model Clustering for Mixed Data with Missing Data, *Computational Statistics & Data Analysis*, 41, 429-440, 2003.
18. İstatistiksel Analize Hazırlık (2): Veritabanında Hata Denetimi, <http://bilgi.umedia.org.tr/yayin/tejm/istkurs2.htm> (Download 13.04.2003).
19. Kneipp S. M., McIntosh M.: Handling Missing Data in Nursing Research with Multiple Imputation, *Nursing Research*, 50, 6, 384-389, 2001.
20. Laurikala J., Juhola M., Lammi S., Penttinen J. Aukee P.: Analysis of Imputed Female Urinary Incontinence Data for the Evaluation of Expert System Parameters, *Comput. Biol. Med.*, 31(4): 239-57, 2001.



## KAYNAKLAR DİZİNİ (Devam)

21. Laurikkala J., Kentala E., Juhola M., Pyykko I.: Treatment of Missing Values with Imputation for the Analysis of Otologic Data, Stud. Health Technol. Inform., 68, 428-31, 1999.
22. Little R. J. A., Rubin D. R.: Statistical Analysis with Missing Data. Wiley, New York, 1987.
23. Little R. J. A.: A Test of Missing Completely at Random for Multivariate Data with Missing Values. Journal of the American Statistical Association 38, 1198-1202, 1998.
24. Little R. J. A., Rubin D. R.: Statistical Analysis with Missing Data, Second Edition, Wiley, New York, 2002.
25. Musil C. M., Warner C. B., Yobas P. K. Jones S. L.: A Comparison of Imputation Techniques for Handling Missing Data, West J. Nurs. Res. 24 (7), 815-29, 2002.
26. Ng H. K. T., Chan P. S., Balakrishnan N.: Estimation of Parameters from Progressively Censored Data Using EM Algorithm, Computational Statistics & Data Analysis, 39, 371-386, 2002.
27. Oğuzlar A.: Alan Araştırmalarında Kayıp Değer Problemi ve Çözüm Önerileri, V. Ulusal Ekonometri ve İstatistik Sempozyumu, Çukurova Üniv. İİBF Ekonometri Bölümü, Adana, 19-22 Eylül 2001.
28. Özdamar K.: Modern Bilimsel Araştırma Yöntemleri, Kaan Kitabevi, Eskişehir, 2003.
29. Özdamar K.: Paket Programlarla İstatistiksel Veri Analizi – 1, Kaan Kitabevi, Eskişehir, 2002.
30. Özdamar K.: Paket Programlarla İstatistiksel Veri Analizi – 2, Kaan Kitabevi, Eskişehir, 2002.

## KAYNAKLAR DİZİNİ (Devam)

31. Özdamar K.: SPSS ile Biyoistatistik, Kaan Kitabevi, Eskişehir, 2003.
32. Patrician P. A.: Multiple Imputation for Missing Data, Res. Nurs. Health, 25 (1), 76-84, 2002.
33. Pigott T. D.: A review of Methods for Missing Data, Educational Research and Evaluation, Vol. 7, No. 4, 353-383, 2001.
34. Sayek İ.: Temel Cerrahi, Cilt 1, Nobel Tıp Kitabevi, 1996.
35. Schafer J. L.: Analysis of Incomplete Multivariate Data, Chapman & Hall, London 1997.
36. Schafer J. L., Olsen M. K.: Multiple Imputation for Multivariate Missing-data Problems: A Data Analyst's Perspective, lecture notes, The Pennsylvania State University, 1998.
37. Schafer J.: Multiple Imputation FAQ Page, <http://www.stat.psu.edu/~jls/mifaq.html>, (Download, 21.04.2002).
38. Schafer J. L.: The Practise of Multiple Imputation, Lecture Notes, The Pennsylvania State University Workshop, March 30, 1998, University Park, PA.
39. Scheffer J.: Dealing with Missing Data, Res. Lett. Inf. Math. Sci, 3, 153-160, 2002.
40. Schimert, J. Schafer, J.L., Hesterberg, T., Fraley, C., Clarkson, D.B., Analyzing Data With Missing Values in S-Plus, Insightful Corp. Seattle, WA. 2001.
41. Schoier G: On Partial Nonresponse Situations: The Hot Deck Imputation Methos, Lecture Notes, Universita di Trieste, Dipartimento di Scienze Economiche e Statistiche, Trieste, Italia, <http://www.stat.fi/isi99/proceedings/arkisto/varasto/scho0502.pdf>. (Download 14.08.2003)
42. Streiner D. L.: The Case of the Missing Data: Methods of Dealing With Dropouts and Other Research Vagaries, Can. J. Psychiatry, Vol. 47, No. 1, 2002.

### **KAYNAKLAR DİZİNİ (Devam)**

43. Sümbüloğlu K., Sümbüloğlu V.: Sağlık Enformasyon Sistemleri, Somgür Eğitim Hizmetleri Yay. Tic. Ltd. Şti, Ankara, 1998.
44. Sümbüloğlu K., Sümbüloğlu V.: Bilimsel Araştırmalarda Biyoistatistik Prensipler ve Yöntemlerinin Bilinçli Kullanımı, Admin Force-Ase Reklamcılık Tesisleri, 2002.
45. Venables W. N., Ripley B. D.: Modern Applied Statistics with S-Plus, Springer-Verlag New York, Inc., 1994.
46. Werner W.: Longitudinal and Multi-group Modelling with Missing Data, White Papers on Structural Equation Modelling, 2000.  
<http://www.smallwaters.com/whitepapers/longmiss/> (Download 11.04.2002).
47. West C. P., Dawson J. D.: Complete Imputation of Missing Repeated Categorical data: One-sample Applications, Stat. Med., 21 (2), 203-17, 2002.
48. Yuan Y.: Multiple Imputation for Missing Data: Concepts and New Development, SAS Institute Inc., <http://www.sas.com/rnd/app/papers/multipleimputation.pdf> (Download 15.06.2003).
49. Zhou X. H., Eckert G. J., Tierney W. M.: Multiple Imputation in Public Health Research, Stat. Med., 20 (9-10), 1541-9, 2001.

## ÖZGEÇMİŞ

1970 yılında Ankara'da doğdu. İlkokulu Akşemsettin İlkokulu'nda 1981 yılında, Ortaokulu Akşemsettin Ortaokulu'nda 1984 yılında, Liseyi Ankara Dikmen Lisesinde 1988 yılında bitirdi. 1988 yılında Anadolu Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü'ne girdi. 1992 yılında bu bölümden mezun oldu. 1992-1993 öğrenim yılında ODTÜ'nde 1 yıl İngilizce hazırlık kursuna devam etti. 1993 yılında Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü'nün Biyoistatistik ABD'nde yüksek lisans öğrenimine başladı. 1995 yılının Ocak ayında Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü'ne Araştırma Görevlisi olarak atandı. Ocak 1997'de yüksek lisans öğrenimini tamamladı. Şubat 1997'de aynı enstitüde Biyoistatistik doktora eğitimine başladı. Halen Osmangazi Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı'nda araştırma görevlisi olarak görevini sürdürmektedir.

Y.C. YÜKSEKÖĞRETİM KURULLU  
DOKÜMANTASYON MERKEZİ