

T.C.
ORDU ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİKTE GEOMETRİNİN UYGULANMASI

EMİNE DEMİRCİ

YÜKSEK LİSANS TEZİ

ORDU 2017

TEZ ONAY

Ordu Üniversitesi Fen Bilimleri Enstitüsü öğrencisi Emine DEMİRCİ tarafından hazırlanan ve Prof. Dr. Cemil YAPAR danışmanlığında yürütülen “ İstatistikte Geometrinin Uygulanması” adlı bu tez, jürimiz tarafından 26 / 05 / 2017 tarihinde oy birliği / oy çokluğu ile Matematik Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman : Prof. Dr. Cemil YAPAR

Başkan : Prof. Dr. Cemil YAPAR
Matematik, Ordu Üniversitesi

İmza :

Üye : Doç. Dr. Selahattin MADEN
Matematik, Ordu Üniversitesi

İmza :

Üye : Yrd. Doç. Dr. Nurgül OKUR BEKAR
İstatistik, Giresun Üniversitesi

İmza :

ONAY:

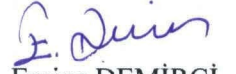
22/06/2017 tarihinde enstitüye teslim edilen bu tezin kabulü, Enstitü Yönetim Kurulu'nun 22/06/2017 tarih ve 2017 / 279 sayılı kararı ile onaylanmıştır.

Enstitü Müdürü
Prof. Dr. Kürşat KORKMAZ



TEZ BİLDİRİMİ

Tez yazım kurallarına uygun olarak hazırlanan bu tezin yazılmasında bilimsel ahlak kurallarına uyulduğunu, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezin içerdiği yenilik ve sonuçların başka bir yerden alınmadığını, kullanılan verilerde herhangi bir tahrifat yapılmadığını, tezin herhangi bir kısmının bu üniversite veya başka bir üniversitedeki başka bir tez çalışması olarak sunulmadığını beyan ederim.


Emine DEMİRÇİ

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ÖZET

İSTATİSTİKTE GEOMETRİNİN UYGULANMASI

Emine DEMİRCİ

Ordu Üniversitesi
Fen Bilimleri Enstitüsü
Matematik Anabilim Dalı, 2017
Yüksek Lisans Tezi, 76s.

Danışman: Prof. Dr. Cemil YAPAR

Bu tez yedi bölümden oluşur. Birinci bölümde tezin amaçları ile ilişkili olarak bir giriş verilir. İkinci bölümde geometrinin istatistik alanında kullanımına dair yapılan çalışmalar hakkında bilgi verilmiştir. Üçüncü bölümde hipotez testleri için geometrik bir bakış ortaya konur. Dördüncü bölümde üç bileşene sahip olan ikili örneklem durumu için bazı geometrik yorumlar vardır. Böylece ele alınan konular daha açık bir hale gelir. Beşinci bölümde t- testi için farklı bir geometrik yaklaşım ele alınır. Altıncı bölüm bağımsız örneklemelerin bir uygulamasını içerir. Son bölümde basit regresyon analizi için bazı bilgiler ve geometrik yorumlar sunulur.

Anahtar Kelimeler: Basit Regresyon, F-Testi, İkili Örneklem, İzdüşüm Uzunlukları Koordinat Sistemi, Ortogonal Ayrışım, t-Testi,

ABSTRACT

SOME APPLICATIONS OF GEOMETRY IN STATISTICS

Emine DEMİRÇİ

University of Ordu
Institute for Graduate Studies in Science and Technology
Department of Mathematics, 2017
MSc. Thesis, 76p.

Supervisor: Prof. Dr. Cemil YAPAR

This thesis consists of seven chapters. In the first chapter, an introduction regarding to the aims of this thesis is given. The second chapter, it is given information about the studies made on the useage of geometric in the field of statistics. The third chapter, a geometric approximation for testing hypothehes is introduced. In the fourth chapter, some geometric interpretations are given for paired sample case having three components. Therefore, treated subjects become clearer. In the fifth chapter, a geometric view for t-test is given. The sixth chapter includes an application of indepented samples. In the last chapter, some information and geometric interpratations for simple regression analysis are presented.

Key Words: Simple Regression, F-Test, Paired Sample, Projection Lengts, Coordinate System, Orthongonal Decomposition, t-Test

TEŐEKKÜR

Tüm alıőmalarım boyunca kıymetli zamanımı ayırarak alıőmalarıma yardımcı olan, engin bilgi ve deneyimleriyle bana yol gösteren deęerli danıőmanım Prof. Dr. Cemil YAPAR'a en iten teőekkürlerimi sunarım.

Ayrıca, her zaman bilgi ve görüőlerine baővurduęum Do. Dr. Selahattin MADEN ve Yrd. Do. Dr. Erdal ÜNLÜYOL ile tüm Ordu Üniversitesi Matematik Bölümü öğretim üyelerine teőekkür ederim.

Hem bu süreçte hem de hayatım boyunca yanımda olan deęerli aileme ve arkadaşlarıma teőekkür ederim.



İÇİNDEKİLER

TEZ BİLDİRİMİ	I
ÖZET.....	II
ABSTRACT.....	III
TEŞEKKÜR	IV
İÇİNDEKİLER.....	V
ŞEKİLLER LİSTESİ.....	VIII
ÇİZELGELER LİSTESİ.....	XI
SİMGELER ve KISALTMALAR.....	XII
EK LİSTESİ.....	XIII
1. GİRİŞ	I
2. ÖNCEKİ ÇALIŞMALAR.....	2
3. HİPOTEZ TESTLERİ İÇİN GEOMETRİK YAKLAŞIM.....	3
3.1. İzdüşüm Uzunluklarının Dağılımı	11
4. n = 3 İKİLİ ÖRNEKLEMLER İÇİN t- TESTİ	14
4.1. Uygun Bir Koordinat Sistemi	15
4.2. Ortogonal Ayrışım	16
4.3. İzdüşüm Uzunluklarının Dağılımları	17
4.4. Hipotezi Test Etme	18
4.5. İkili Örneklem Testi.....	19
4.6. Yorumlar	20
5. t- TESTİNE GEOMETRİK BİR BAKIŞ	21
5.1. Geometrik Yol	21
5.2. Gemoetrik ve Geleneksel Yollar Arasındaki Bağlantı.....	25
5.3. θ' dan t- ye Dönüşüm.....	26
5.4. $H_0 : \mu = \mu_0$ Durumu.....	27

6.	BAĞIMSIZ ÖRNEKLEMLER	28
6.1.	Veri Kümesi	28
6.2.	İki Büyüklüğüne Sahip Örneklemeler	30
6.2.1.	Amaç	30
6.2.2.	Temel Düşünce	30
6.2.3.	Uygun Bir Koordinat Sistemi	31
6.2.4.	Ortogonal Ayrışım	32
6.2.5.	İzdüşüm Uzunluklarının Dağılımları	34
6.2.6.	Hipotezi Test Etme	36
6.2.7.	Bağımsız Örneklemeler <i>t</i> -Testi	38
6.2.8.	Yorumlar	39
7.	BASİT REGRESYON	40
7.1.	Hava Kirliliği	40
7.2.	Beş Büyüklüğüne Sahip Örneklemeler	42
7.2.1.	Amaç	44
7.2.2.	Model	44
7.2.3.	Temel Düşünce	45
7.2.4.	Uygun Bir Koordinat Sistemi	46
7.2.5.	Ortogonal Ayrışım	48
7.2.6.	İzdüşüm Uzunluklarının Dağılımları	52
7.2.7.	Hipotezi Test Etme	54
7.2.8.	Eşdeğer <i>t</i> -Testi	56
7.2.9.	Yorumlar	56
7.3.	Genel Durum	57
7.3.1.	Temel Düşünce	57
7.3.2.	Uygun Bir Koordinat Sistemi	58
7.3.3.	Ortogonal Ayrışım	58
7.3.4.	İzdüşüm Uzunluklarının Dağılımları	61

7.3.5.	Hipotezi Test Etme	62
7.3.6.	Denk t -Testi	63
7.3.7.	Korelasyon Katsayısı	64
7.2.8.	Yorumlar	65
7.2.9.	Özet.....	66
8.	SONUÇ ve ÖNERİLER.....	69
9.	KAYNAKLAR	70
	EKLER.....	72
	ÖZGEÇMİŞ	76



ŞEKİLLER LİSTESİ

<u>Sekil No</u>		<u>Sayfa</u>
Şekil 3.1.	$\mu = 0$ durumunda hassaslık ölçümleri	3
Şekil 3.2.	$\mu = 0$ durumunda ölçümlerin y_1 - ve y_2 - eksenli koordinat sistemindeki görünümü	4
Şekil 3.3.	$\mu = 0$ durumunda ölçümlerin yayılımı.....	4
Şekil 3.4.	$\mu \neq 0$ durumunda hassaslık ölçümleri	5
Şekil 3.5.	$\mu \neq 0$ durumunda ölçümlerin y_1 - ve y_2 - eksenli koordinat sistemindeki görünümü	5
Şekil 3.6.	$\mu \neq 0$ durumunda ölçümlerin yayılımı	6
Şekil 3.7.	$\mu = 0$ durumunda A ve B uzaklıkları	7
Şekil 3.8.	$\mu \neq 0$ durumunda A ve B uzaklıkları	8
Şekil 3.9.	(y_1, y_2) durumunda A ve B uzaklıkları	8
Şekil 3.10.	Dik koordinat sistemi.....	9
Şekil 3.11.	Test istatistiğinin hesaplanması (t).....	10
Şekil 3.12.	Test istatistiğinin hesaplanması (F).....	11
Şekil 3.13.	$y \cdot U_1$ ve $y \cdot U_2$ izdüşüm uzunluklarının dağılımları.....	13
Şekil 4.1.	İncelememizin birçok tekrarını temsil eden veri noktalarının dağılımları.....	15
Şekil 4.2.	$(y - \bar{y})$ cinsinden gözlem vektörünün ortogonal ayrışımı	17
Şekil 4.3.	Üç ortogonal izdüşümün uzunluklarının karelerinin Pisagor parçalanması.....	19
Şekil 5.1.	(a) $\mu = 0$, (b) $\mu \neq 0$ ve (c) $\mu \neq 0$ durumlarında bir incelemenin birçok tekrarından ortaya çıkan veri vektörleri	22

Şekil 5.2.	$r_0 = \sqrt{(19.0)^2 + (7.6)^2 + (22.9)^2} = 30.711$ yarıçaplı küreye ilişkin $(19.0, 7.6, 22.9)'$ veri vektörü.....	24
Şekil 5.3.	Veri vektörünün vektör ayrışımı	26
Şekil 5.4.	$t = \sqrt{2} \cot(\theta)$ dönüşümü	27
Şekil 6.1.	Kalp atım hızını temsil eden veri noktalarının bulutları	31
Şekil 6.2.	İki inceleme kitlesinden alınan iki büyüklüğüne sahip bağımsız örneklem durumu için gözlem vektörünün ortogonal ayrışımı.....	34
Şekil 6.3.	$\mathbf{y} \cdot \mathbf{U}_1, \mathbf{y} \cdot \mathbf{U}_2, \mathbf{y} \cdot \mathbf{U}_3$ ve $\mathbf{y} \cdot \mathbf{U}_4$ izdüşüm uzunluklarının dağılımlarının resimsel gösterimi.....	36
Şekil 6.4.	Gözlem vektörünün uzunluğunun karesinin Pisagor parçalanması.....	37
Şekil 7.1.	1992 kışı haziran ayı için inversiyon etkisine karşı hava kirliliği seviyesinin saçılım diyagramı	42
Şekil 7.2.	5 günlük örneklem için inversiyon etkisine karşı hava kirliliği seviyesinin saçılım diyagramı	43
Şekil 7.3.	Basit regresyon modelinin varsayımları	44
Şekil 7.4.	Hava kirliliği incelememizin birçok tekrarlanmasını temsil eden veri noktalarının dağılımları	46
Şekil 7.5.	\mathbf{U}_1 ve \mathbf{x} vektörü tarafından tanımlanan düzlemde $(\mathbf{x} - \bar{\mathbf{x}})$ doğrultusu...	47
Şekil 7.6.	Basit regresyon durumu için gözlem vektörünün ortogonal ayrışımı.....	50
Şekil 7.7.	İnversiyon etkisine karşı hava kirliliğinin saçılım diyagramı üzerinde birleştirilmiş uygun regresyon.....	51
Şekil 7.8.	Gözlem vektörünün uzunluğunun karesinin, beş ortogonal izdüşümün kareleri toplamı olarak Pisagor parçalanması	55
Şekil 7.9.	Kirlilik incelememizin birçok tekrarlanmasının temsil eden veri noktalarının dağılımları	58
Şekil 7.10.	Çizelge 6.1'den tam veri kümesini kullanan, basit regresyon için gözlem vektörünün ortogonal ayrışımı	60
Şekil 7.11.	İnversiyon etkisine karşı hava kirliliğinin saçılım diyagramı üzerine yerleştirilen uygun regresyon doğrusu	61

Şekil 7.12.	Basit regresyon için gözlem vektörünün uzunluğunun karesinin Pisagor parçalanması	62
Şekil 7.13.	$(y - \bar{y})$ düzeltilmiş gözlem vektörü ve kosinüsü r korelasyon katsayısı olan θ açısını gösteren ortogonal ayrışım	64
Şekil 7.14.	Logaritmaya dönüştürülen hava kirliliği ile yeniden gösterimi.....	66
Şekil 7.15.	Basit regresyon durumu için gözlem vektörünün ortogonal ayrışımı.....	67



ÇİZELGELER LİSTESİ

<u>Çizelge No</u>		<u>Sayfa</u>
Çizelge 4.1.	26 tarım araştırma görevlisi için 30 saniyede kalp atış sayısındaki değişmeler	14
Çizelge 5.1.	Kadın ve erkeklerden oluşan 3 örneklem	21
Çizelge 6.1.	Bölüm 3'teki 26 tarım araştırmacısının koşudan önce ve sonraki gerçek kalp atışları	29
Çizelge 6.2.	İki dingin ve iki koşu sonrası kişinin örneklem ortalaması ve kalp atım hızları.....	30
Çizelge 7.1.	Hava kirliliği seviyeleri, yer seviyesinde ve yerin yukarısında sıcaklıklar ve onların farkı(inversiyon etkisi).....	41
Çizelge 7.2.	Başlangıç ile ilgili örneğimiz için kullanılan 5 günlük örneklem.....	43

SİMGE ve KISALTMALAR

α : regresyon sabiti

p : olasılık

r_0 : yarıçap

$f(r)$: olasılık yoğunluk fonksiyonu

\cdot : iç çarpım

β : eğim

μ : ortalama

σ : varyans

EK LİSTESİ

EK No

Sayfa

EK 1. İstatistiksel Tablolar73



GİRİŞ

“Bin söz söyleyeceğine bir şekil yap göster.” özdeyişinden hareketle, istatistiksel kavramların geometrik yorumu onların daha kolay anlaşılmasını sağlar. Bir geometrik şekil, kısa bir bakışta birçok düşünce uyandırarak problem ve çözümü arasında bir köprü oluşturabilir. Bu nedenle geometrik yorumlar ve çizimler bir istatistiksel problemin daha kolay anlaşılması için farklı bakış ve yöntemleri ortaya koyar.

İstatistiksel kavramlara geometrik yaklaşımlar; klasik yaklaşım ve matris yaklaşımları arasındaki boşluğu doldurur. Bundan dolayı, istatistiksel yöntemler ve onların geometrik yorumları son zamanlarda artan bir biçimde istatistik literatürüne girmiştir. Bu konuda yazılan kitaplar ve makaleler bu bakışın önemini artan bir biçimde vurgulamaktadır.

2. ÖNCEKİ ÇALIŞMALAR

Varyans ve regresyon çözümlenmesi yöntemleri 20.yüzyılın başlarında analitik geometri kullanılarak geliştirilmiştir. Bu alandaki ilk çalışmalar R.A. Fisher (1915) ile başlamıştır. Fisher, geometrik yaklaşımları kullanarak bu yöntemlerin görselleşmesini de sağlamıştır. Geometrik yaklaşımların algılanmasındaki zorluklar sebebiyle uzun bir süre bu yaklaşımların kullanılmasından uzak durulmuştur.

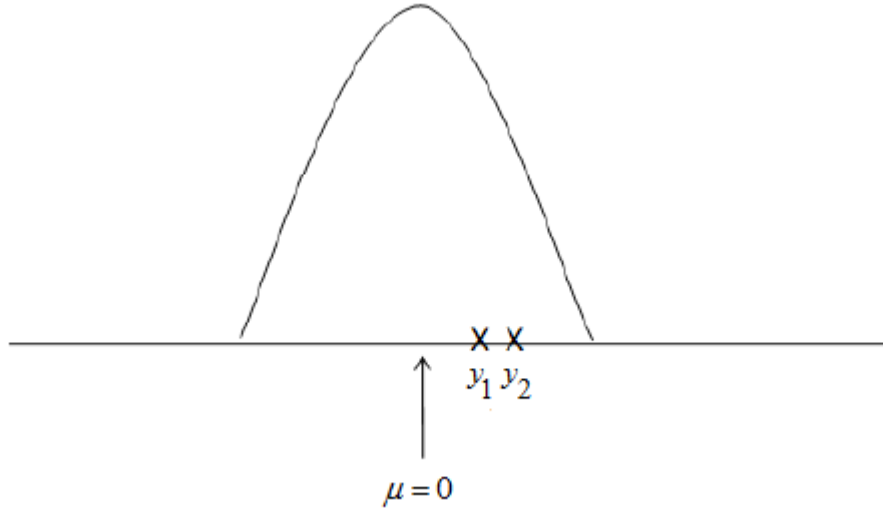
Geometrik yaklaşımlar, Fisher (1915)'in, korelasyon katsayısının dağılımı ile ilgili çalışması ile başlamış, Kruskal (1975)'in genelleştirilmiş ters (inverse) geometrisi hakkında yaptığı çalışma ile devam etmiştir. Doğrusal modellere geometrik yaklaşımlar Fisher (1915), Bartlett (1933-1934), Durbin ve Kendall (1951), Kruskal (1961, 1968, 1975), Zyskind (1967), Watson (1967) ve Saville ve Wood (1983, 1984, 1986, 1990) tarafından incelenmiştir. Demirhan ve Parlak (2002) ise çok etkenli deneysel yöntemlere geometrik yaklaşımları kullanarak bir uygulama yapmışlardır.

Saville ve Wood tarafından yazılan *Statistical Methods: The Geometric Approach* ile *Statistical Methods A Geometric Primer* adlı kitaplar ve yine aynı yazarlar tarafından yazılan bu konudaki birçok makale bu bakışı destekleyen yayınlar arasında, önemli bir yer tutmaktadır. Bu tezde bu ve benzeri yayınlardan önemli derecede faydalanılmaya çalışılmıştır.

3. HİPOTEZ TESTLERİ İÇİN GEOMETRİK YAKLAŞIM

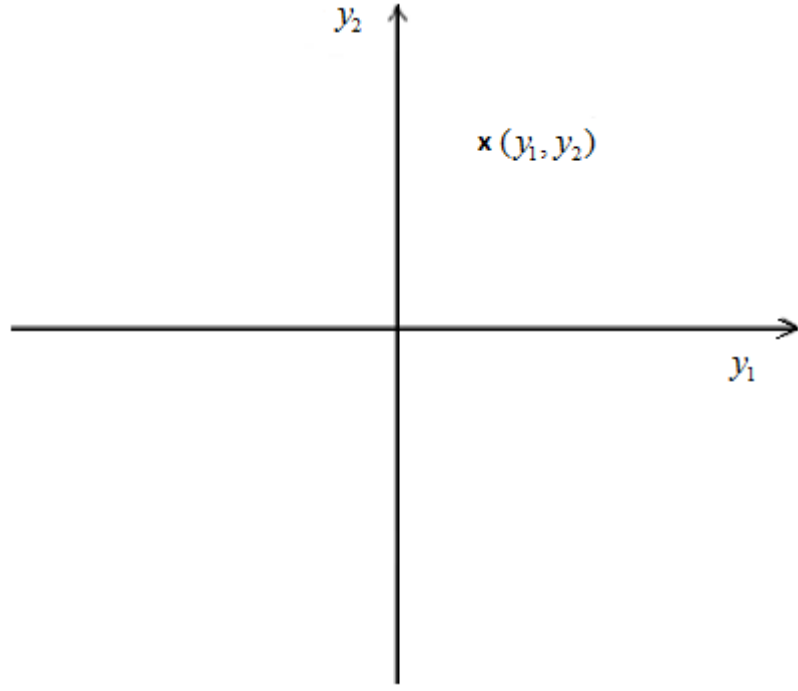
Yaklaşımımızın ardındaki temel düşünceyi açıklamak için en basit mümkün olabilir durumu, yani bir tek kitleden iki büyüklüğüne sahip bir örneklem durumunu ele alalım. Yeni alınan bir terazinin hassas olup olmadığını araştırmak istediğimizi varsayalım. Terazideki ölçümlerde herhangi bir hata olup olmadığını test etmenin bir yöntemi yeterli aralıklarla ölçüm yapıp, oluşan değerleri okumak olmalıdır. Böyle bir incelemeden elde edilen iki veri y_1 ve y_2 ağırlıkları olsun. Terazinin hassas(yanlısapmalı) olup olmadığına karar vermek için bu verileri nasıl kullanabiliriz?

Terazide birçok ölçüm yaptığımızı ve terazinin tam olarak hatasız olduğunu, bu nedenle uygun vadede ağırlık ölçümlerinin $\mu = 0$ ortalamaya yöneldiğini düşünelim. Bu durumda y_1, y_2 Şekil 3.1’de görüldüğü gibi işaretlenebilir.



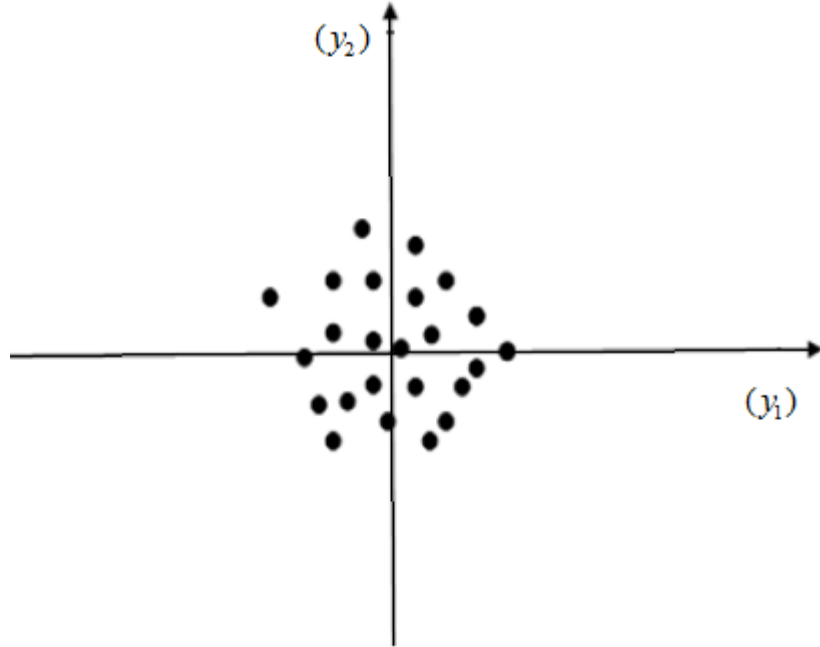
Şekil 3.1. $\mu = 0$ durumunda hassaslık ölçümleri

y_1 ve y_2 'nin $y_1 -$ ve $y_2 -$ eksenlerine sahip bir koordinat sistemindeki görünümüne bakacak olursak Şekil 3.2 elde edilir.



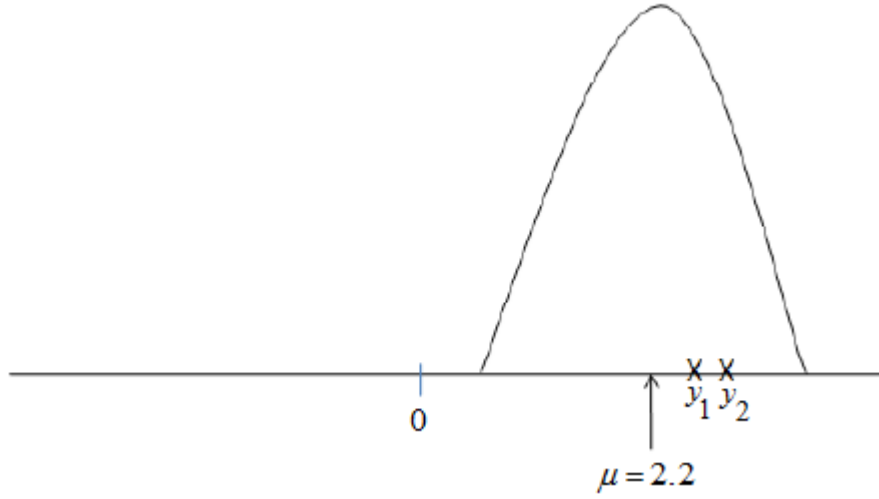
Şekil 3.2. $\mu = 0$ durumunda ölçümlerin y_1 - ve y_2 - eksenli koordinat sistemindeki görünümü

Kitle ortalaması $\mu = 0$ ise, bu takdirde ölçümler orijin etrafında bir yayılım göstermektedir. Bu durum Şekil 3.3'te görülmektedir.

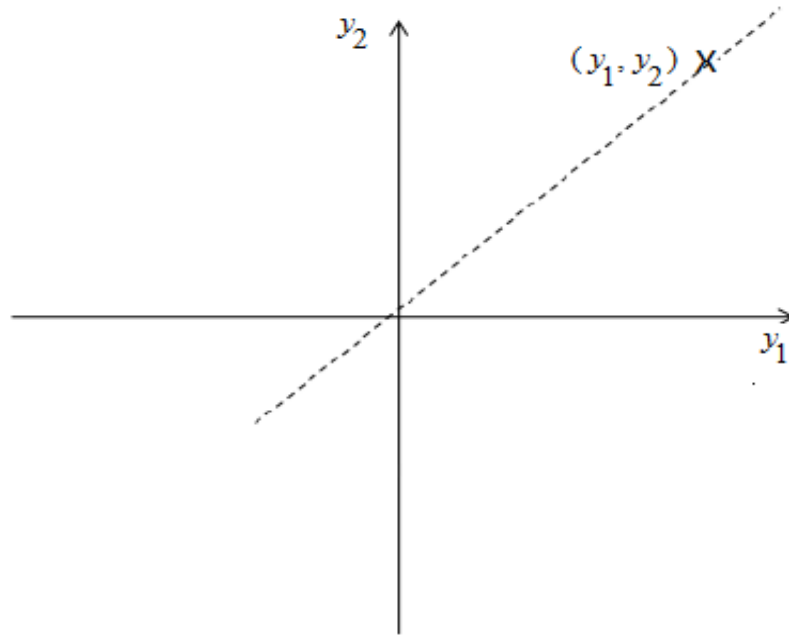


Şekil 3.3. $\mu = 0$ durumunda ölçümlerin yayılımı

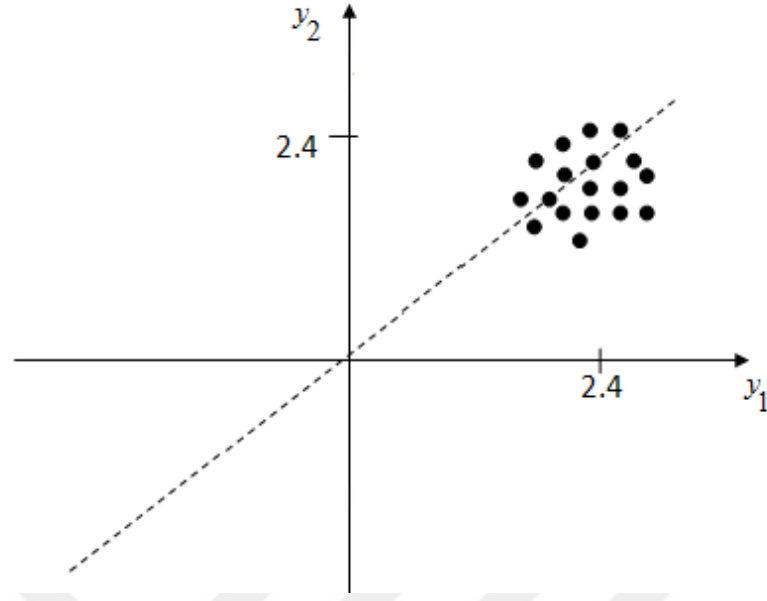
Eğer ölçümler arasında fark varsa; $\mu \neq 0$ olacaktır ve yayılım eksenler ile 45° lik açı yapan bir doğru boyunca kayacaktır. Şekil 3.4, Şekil 3.5 ve Şekil 3.6 bu durumu belirten çizimlerdir.



Şekil 3.4. $\mu \neq 0$ durumunda hassaslık ölçümleri



Şekil 3.5. $\mu \neq 0$ durumunda ölçümlerin y_1 - ve y_2 - eksenli koordinat sistemindeki görünümü

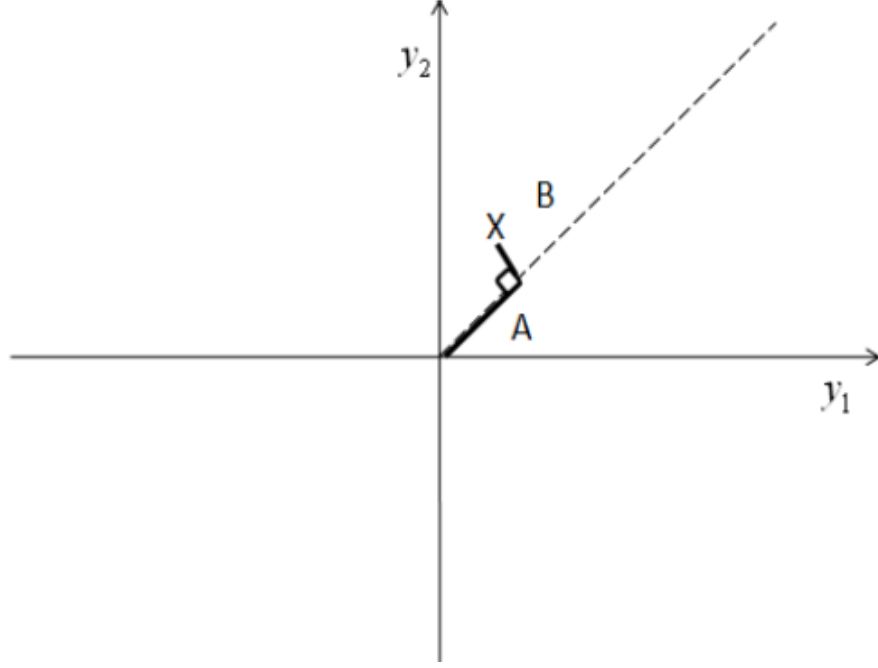


Şekil 3.6. $\mu \neq 0$ durumunda ölçümlerin yayılımı

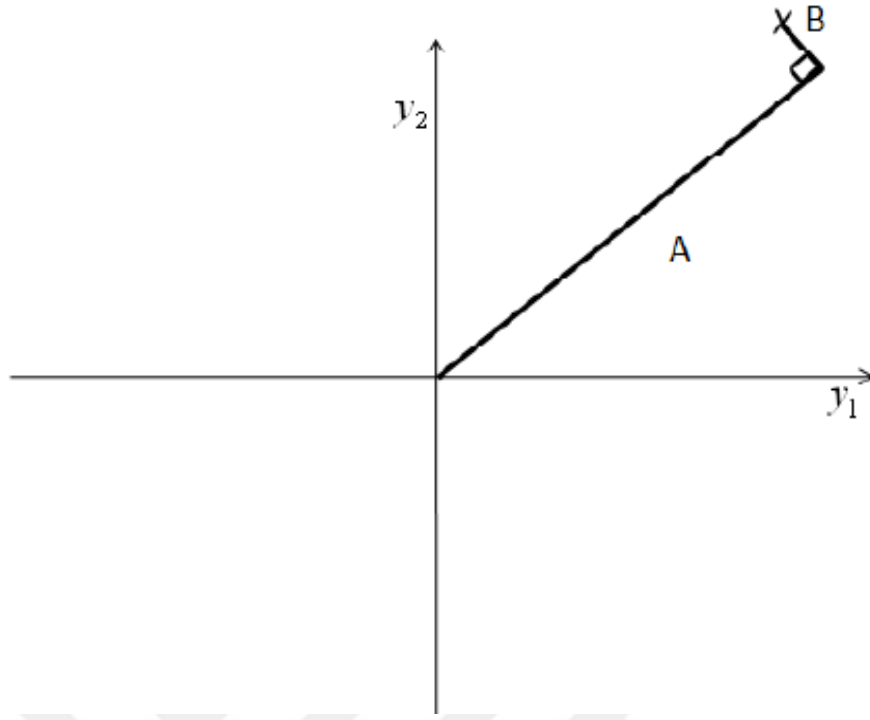
Şimdi tipik bir (\mathbf{x}) örneklem noktası $(\mu, \mu) = (2.4, 2.4)$ noktası üzerinde merkezleşen noktaların saçılımındaki bir noktadır. Terazimiz ölçme konusunda yanlı mıdır? sorumuzu geometrik bir soruya çevirebiliriz. Bir gerçek durumda, $(y_1, y_2) = (2.3, 2.5)$ diyeceğimiz sadece deneysel sonuçlar kümesine sahip olacaksınız. Eşdeğer geometrik soru: Nokta orijinde merkezleşen bir saçılıma mı yoksa orijinden uzakta merkezleşen bir saçılıma mı aittir? Daha kesin olarak, $\mu = 0$ mıdır veya $\mu \neq 0$ mıdır? İhtiyacımız olan şey iki durumu ayıracak olan “test istatistiği” olarak adlandırılan bir ölçüdür. Bu ölçü, eğer $\mu = 0$ ise, bu takdirde küçük ve eğer $\mu \neq 0$ ise, bu takdirde büyük olmalıdır.

Şekil 3.3 ve Şekil 3.6 istediğimiz ipuçlarını verir. $\mu \neq 0$ durumunda, noktaların dağılımı eş açısız doğrunun, yani her iki eksenle 45° lik açığa sahip doğrunun, yukarısında yer alır. “ \mathbf{x} ” gibi bir örneklem noktasının eş açısız doğruya olan uzaklığı B ve bu doğrunun orijine olan uzaklığı A olarak gösterilsin. Bu durumda $\mu \neq 0$ iken elde edilen A/B oranı, $\mu = 0$ durumundakinden daha büyük olur ve A/B oranı test istatistiği olarak kullanılabilir. Bu test istatistiğinin noktaların dağılımı nasıl olursa olsun çalıştığına ve ölçüm birimlerimize bağlı olmadığına dikkat edelim. Bu oranın t-dağılımına sahip olduğu söylenebilir (Saville ve Wood, 1996).

$(y_1, y_2) = (2.3, 2.5)$ örneklem noktamız için test istatistiği Şekil 3.7 ve Şekil 3.8’de gösterildiği gibi, $A/B = 2.4\sqrt{2} / (0.1\sqrt{2}) = 24$ değerini alır. Eğer μ ortalaması gerçekten sıfır ise, bu takdirde test istatistiğimizin 1 serbestlik dereceli Student t dağılımına sahip olduğunu söyleyebiliriz. Bu dağılımın bilgisi bize bir test istatistiğinin büyük olması hakkında kesin olma olanağı verir.

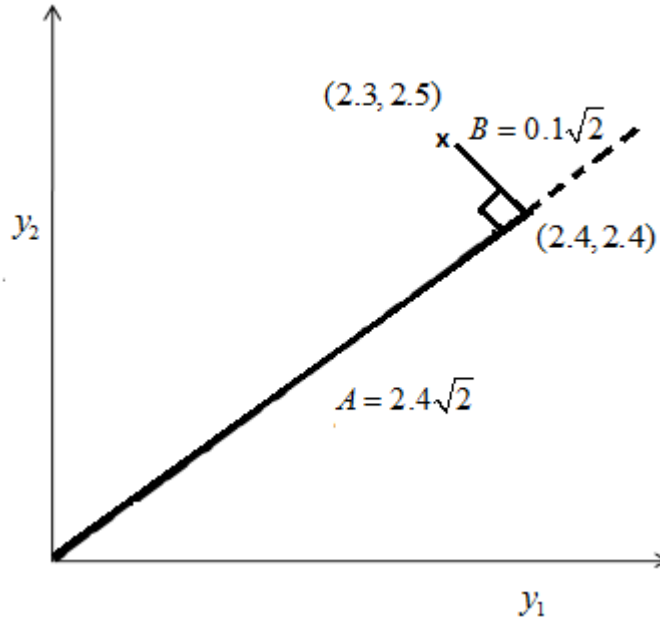


Şekil 3.7. $\mu = 0$ durumunda **A** ve **B** uzaklıkları



Şekil 3.8. $\mu \neq 0$ durumunda **A** ve **B** uzaklıkları

Örneğimizde $(y_1, y_2) = (2.3, 2.5)$ için bu oran Şekil 3.9'da gösterilmiştir.

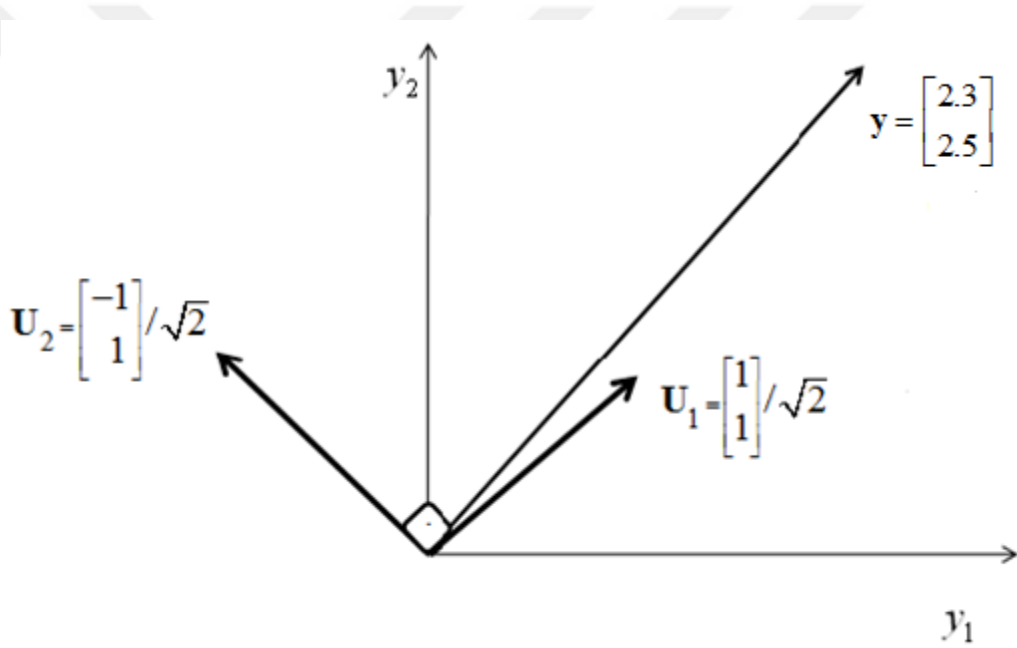


Şekil 3.9. (y_1, y_2) noktası ve **A** ve **B** uzaklıkları

Örneğin, t_1 istatistiğinin mutlak değeri durumların %95'inde 12.7'den küçüktür. Geleneksel olarak, istatistikçiler bu değerlerin büyük beyan edilen 12.7'den daha büyük değerlere göre, küçük olacağını iddia eder. Bu, 24'lük değerimizin büyük

iddia edilmesi gerektiğini, bu nedenle terazinin yanlı olduğu düşüncesini desteklemek için yeterli delile sahip olduğumuzu ifade eder.

Benzer analiz vektörler ile yapılmak istendiğinde ilk olarak; iki boyutlu uzayda gözlemleri noktalar yerine vektörlerle ifade ederiz. Bu durumda gözlemimizi $\mathbf{y} = [2.3, 2.5]'$ vektörü olarak gösteririz. İkinci olarak; iki boyutlu uzay için bir dik koordinat sistemine ihtiyaç duyulur. Eşit açılı bir yön elde edebilmek için, ilk koordinat eksenini olarak $\mathbf{U}_1 = [1, 1]'/\sqrt{2}$ seçilirse, bu takdirde $\mathbf{U}_2 = [-1, 1]'/\sqrt{2}$ 'de diğer koordinat eksenini olarak seçilmiş olur. Bu durum Şekil 3.10'da görselleştirilmiştir.

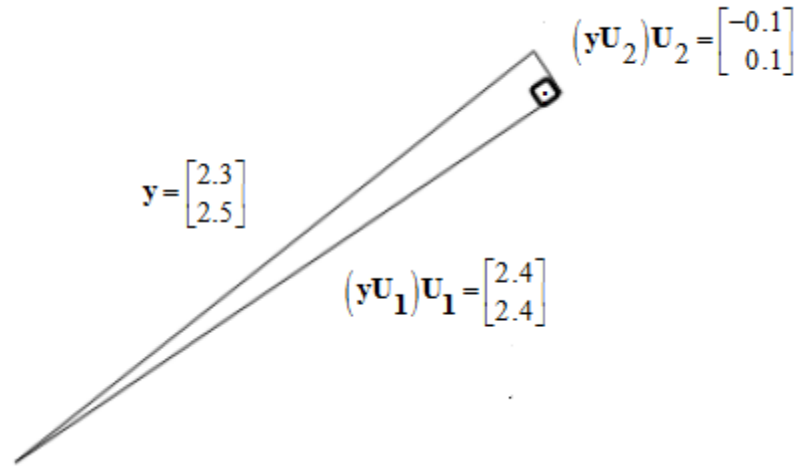


Şekil 3.10. Dik koordinat sistemi

Üçüncü olarak, \mathbf{y} gözlem vektörümüzü sırasıyla koordinat eksenlerimizin her biri üzerine izdüşürürüz ve Şekil 3.11'de gösterildiği gibi

$$(\mathbf{y}\mathbf{U}_1)\mathbf{U}_1 = 2.4\sqrt{2}\mathbf{U}_1 = \begin{bmatrix} 2.4 \\ 2.4 \end{bmatrix} \text{ ve } (\mathbf{y}\mathbf{U}_2)\mathbf{U}_2 = 0.1\sqrt{2}\mathbf{U}_2 = \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix} \quad (3.1)$$

izdüşüm vektörlerini elde ederiz.



Şekil 3.11. Test istatistiğinin hesaplanması (t)

Bu işlem, gözlem vektörünün biri “model vektörü” ve diğeri “hata vektörü” olarak iki vektöre ortogonal ayrışımını ortaya koyar. Bu ortogonal ayrışım (3.2) bağıntısında gösterilmiştir.

$$\mathbf{y} = (\mathbf{y}\mathbf{U}_1)\mathbf{U}_1 + (\mathbf{y}\mathbf{U}_2)\mathbf{U}_2 \quad (3.2)$$

Gözlem vektörü = Model vektörü + Hata vektörü

Örneğe dönecek olursak,

$$\mathbf{y} = 2.4\sqrt{2}\mathbf{U}_1 + 0.1\sqrt{2}\mathbf{U}_2 \quad (3.3)$$

$$\begin{bmatrix} 2.3 \\ 2.5 \end{bmatrix} = \begin{bmatrix} 2.4 \\ 2.4 \end{bmatrix} + \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix} \quad (3.4)$$

elde edilir. Bu, $|\mathbf{y}\mathbf{U}_2|$; $\mathbf{y}\mathbf{U}_2$ 'nin büyüklüğünü ifade etmek üzere,

$$\frac{|\mathbf{y}\mathbf{U}_1|}{|\mathbf{y}\mathbf{U}_2|} = \frac{2.4\sqrt{2}}{0.1\sqrt{2}} = 24 \quad (3.5)$$

test istatistiğine götürür. Bu test istatistiğinin önceki kısımda elde edilen A/B test istatistiği olduğuna dikkat ediniz.

t dağılımına ek olarak F dağılımı kullanılmak istenirse; Pisagor teoreminden yararlanarak gözlem vektörünün 2 boyutlu uzaydaki uzunluğunun karesi;

$$\|\mathbf{y}\|^2 = (\mathbf{y}\mathbf{U}_1)^2 + (\mathbf{y}\mathbf{U}_2)^2 \quad (3.6)$$

şeklinde bulunur. Buradan yararlanarak,

$$F = \frac{(\mathbf{y} \cdot \mathbf{U}_1)^2}{(\mathbf{y} \cdot \mathbf{U}_2)^2} = \frac{A^2}{B^2} \quad (3.7)$$

F dağılımına sahip bir test istatistiği daha elde edebiliriz. Örneğimizde,

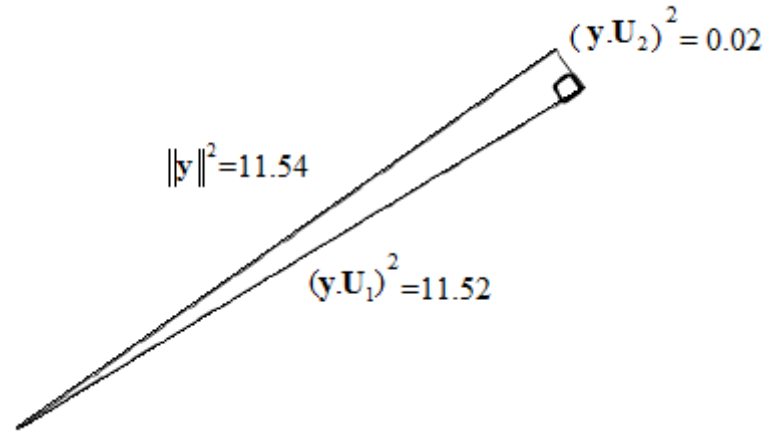
$$2.3^2 + 2.5^2 = (2.4\sqrt{2})^2 + (0.1\sqrt{2})^2 \quad (3.8)$$

$$11.54 = 11.52 + 0.02 \quad (3.9)$$

ve

$$F = \frac{(\mathbf{y} \cdot \mathbf{U}_1)^2}{(\mathbf{y} \cdot \mathbf{U}_2)^2} = \frac{11.52}{0.02} = 576 \quad (3.10)$$

olarak elde edilir. Bu durum aşağıda verilen Şekil 3.12 üzerinde de görülmektedir. Bu değer önceki kısımdaki notasyon cinsinden A^2 / B^2 olduğuna dikkat ediniz. Bu değer $F_{1,1}$ dağılımının yüzdeleri ile (örneğin 161,95) karşılaştırılır ve yine bizi terazinin yanlı olduğu fikrini desteklemek için yeterli delile sahip olduğumuz sonucunu çıkarmaya sevk eder.



Şekil 3.12. Test istatistiğinin hesaplanması (F)

3.1. İzdüşüm Uzunluklarının Dağılımı

$\mathbf{y} \cdot \mathbf{U}_1$ ve $\mathbf{y} \cdot \mathbf{U}_2$ izdüşüm uzunluklarının dağılımını inceleyecek olursak; y_1 ve y_2 sembolleri, μ ortalamalı ve σ^2 varyanslı bir normal dağılımdan gelen bağımsız gözlemleri ifade etmek üzere,

$$\mathbf{y} \cdot \mathbf{U}_1 = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{y_1 + y_2}{\sqrt{2}} \quad (3.11)$$

eşitliğinden yararlanarak $\mathbf{y} \cdot \mathbf{U}_1$ izdüşümünün ortalamasını,

$$\mathbf{y} \cdot \mathbf{U}_1 = \frac{\mu + \mu}{\sqrt{2}} = \frac{2\mu}{\sqrt{2}} = \sqrt{2}\mu \quad (3.12)$$

ve varyansını,

$$\mathbf{y} \cdot \mathbf{U}_1 = \left(\frac{1}{\sqrt{2}} \right)^2 [\text{varyans}(y_1) + \text{varyans}(y_2)] = \frac{1}{2} [\sigma^2 + \sigma^2] = \sigma^2 \quad (3.13)$$

şeklinde elde ederiz. Bu nedenle, $\mathbf{y} \cdot \mathbf{U}_1$ izdüşümü; $\sqrt{2}\mu$ ortalamalı ve σ^2 varyanslı bir normal dağılımdan yani, bir $N(\sqrt{2}\mu, \sigma^2)$ dağılımından gelir. Aynı şekilde $\mathbf{y} \cdot \mathbf{U}_2$ izdüşümünün dağılımını elde edebilmek için,

$$\mathbf{y} \cdot \mathbf{U}_2 = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \frac{y_2 - y_1}{\sqrt{2}} \quad (3.14)$$

eşitliğinden yararlanarak $\mathbf{y} \cdot \mathbf{U}_2$ izdüşümünün ortalamasını,

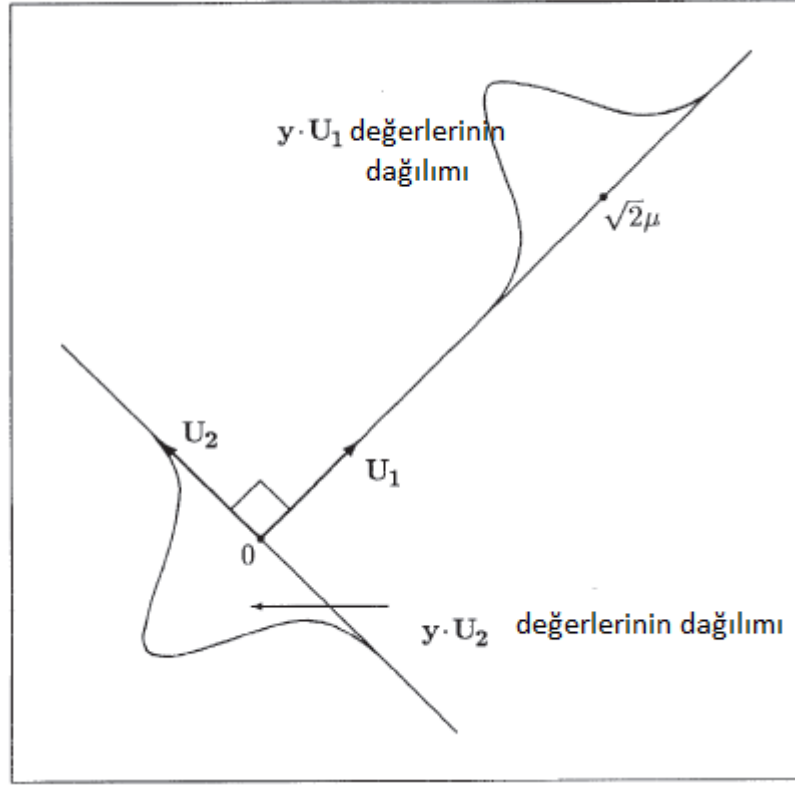
$$\mathbf{y} \cdot \mathbf{U}_2 = \frac{\mu - \mu}{\sqrt{2}} = 0 \quad (3.15)$$

ve varyansını,

$$\mathbf{y} \cdot \mathbf{U}_2 = \left(\frac{1}{\sqrt{2}} \right)^2 [\sigma^2 + \sigma^2] = \sigma^2 \quad (3.16)$$

şeklinde elde ederiz. Bu nedenle $\mathbf{y} \cdot \mathbf{U}_2$ izdüşümü; 0 ortalamalı, σ^2 varyanslı bir normal dağılımdan, yani bir $N(0, \sigma^2)$ dağılımından gelir.

$\mathbf{y} \cdot \mathbf{U}_1$ ve $\mathbf{y} \cdot \mathbf{U}_2$ 'nin dağılımları Şekil 3.13'te gösterilmiştir. Dikkat edilmesi gereken nokta, $\mathbf{y} \cdot \mathbf{U}_2$ 'nin dağılımı daima sıfırda merkezileşirken, $\mathbf{y} \cdot \mathbf{U}_1$ 'in dağılımının potansiyel olarak sıfırdan farklı bir nicelik(sayı) üzerinde merkezileşmiş olmasıdır.



Şekil 3.13. $y \cdot U_1$ ve $y \cdot U_2$ izdüşüm uzunluklarının dağılımları

4. n=3 İKİLİ ÖRNEKLEMLER İÇİN t- TESTİ

Üç büyüklüğüne sahip bir örneklem için farklı olarak ne yapabiliriz? Şimdi asıl fark, σ^2 kitle varyansının tahmini için mevcut iki koordinat eksenini doğrultularına sahip olmamızdır. Bunun yöntemimizi nasıl etkilediğini göstermek için Çizelge 4.1'deki 7, 18 ve 23 numaralı şahıslardan oluşan rasgele örneklemi yeni örneğimiz olarak kullanarak, Bölüm 3'ü tekrar çalışırız.

Kalp atışı sayısının egzersizle artıp artmadığını belirlemek için 26 tarım araştırmacısı bir deneye tabii tutuldu. Her bir şahıs 10 dakika oturduktan sonra ve yine hemen orada 2 dakika koştuğundan sonra kalp atışı sayısı alındı. Her iki olayda da 30 saniye boyunca kalp atışlarının sayısı kaydedildi. 30 saniyede kalp atışlarındaki değişme Çizelge 4.1'de verilmiştir (Saville ve Wood, 1996).

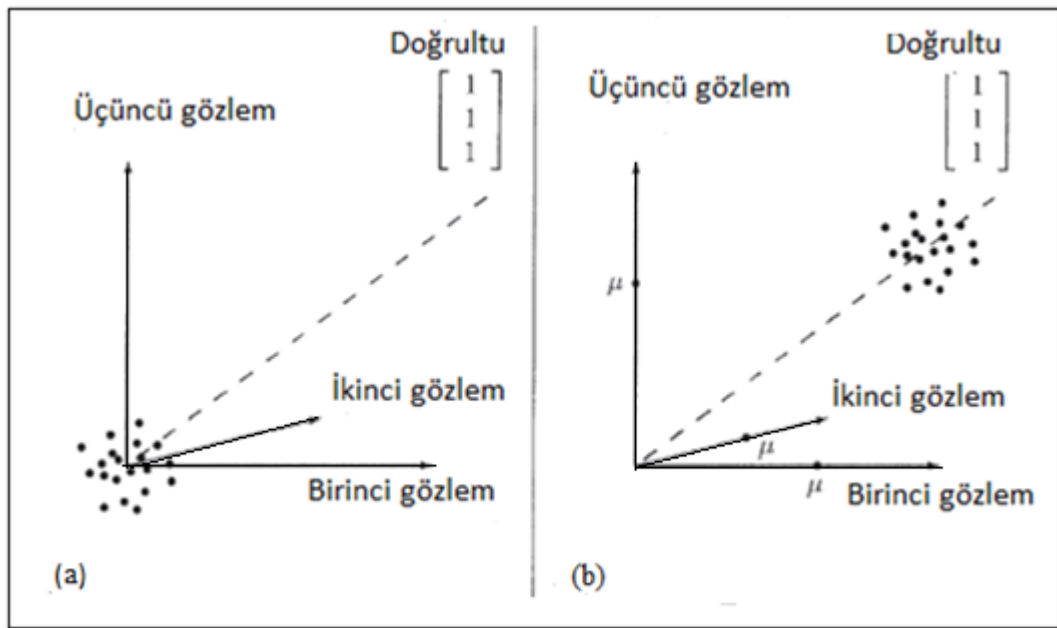
Çizelge 4.1. 26 tarım araştırma görevlisi için 30 saniyede kalp atış sayısındaki değişmeler

Şahıs	Değişme	Şahıs	Değişme
1	7	14	8
2	10	15	10
3	14	16	17
4	0	17	7
5	13	18	14
6	5	19	5
7	7	20	4
8	7	21	16
9	6	22	12
10	20	23	9
11	9	24	0
12	8	25	12
13	13	26	13

Şahısların kalp atım hızı arasındaki farklar, sırasıyla 7, 14, 9'dur. Elde edilen gözlem vektörü

$$\mathbf{y} = \begin{bmatrix} 7 \\ 14 \\ 9 \end{bmatrix} \quad (4.1)$$

dur. Eğer kalp atım hızıyla ilgili μ ortalama ağırlığı sıfır ise, bu takdirde $(7,14,9)'$ veri noktamız Şekil 4.1(a)'da gösterilen dağılımın bir parçası olacak, eğer μ sıfır değilse bu veri noktamız Şekil 4.1(b)'deki dağılımın bir parçası olacak. $\mu = 0$ ve $\mu \neq 0$ olabilirlikleri arasında karar vermek için yöntemimiz; $(7,14,9)$ veri noktamızı belirleyen vektörün eş açılı doğrultuya (\mathbf{U}_1 'e) dik izdüşüm vektörünün uzunluğunun karesini, \mathbf{y} 'nin, \mathbf{U}_1 'e dik olan diğer \mathbf{U}_2 ve \mathbf{U}_3 vektörleri üzerine dik izdüşümlerinin uzunlukları kareleri ortalaması ile karşılaştırmaktır.



Şekil 4.1. İncelememizin birçok tekrarının temsil eden veri noktalarının dağılımları

4.1. Uygun Bir Koordinat Sistemi

Amacımıza özel 3-boyutlu uzay için ortogonal koordinat eksenlerinin bir kümesi

$$\mathbf{U}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{U}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{U}_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \quad (4.2)$$

dır. Burada $\mathbf{U}_1 = [1, 1, 1]' / \sqrt{3}$ ilgilenilen μ parametresi ile ilgili doğrultudur.

4.2. Ortogonal Ayrışım

Şimdi $\mathbf{y} = [7, 14, 9]'$ gözlem vektörümüzü koordinat eksenlerimizin her biri üzerine izdüşürürüz. Birinci izdüşüm vektörü

$$(\mathbf{y} \cdot \mathbf{U}_1) \mathbf{U}_1 = \left(\begin{bmatrix} 7 \\ 14 \\ 9 \end{bmatrix} \cdot \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right) \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 10 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad (4.3)$$

dir. İkinci izdüşüm vektörü

$$(\mathbf{y} \cdot \mathbf{U}_2) \mathbf{U}_2 = \left(\begin{bmatrix} 7 \\ 14 \\ 9 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right) \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -3.5 \\ 3.5 \\ 0 \end{bmatrix} \quad (4.4)$$

dır. Üçüncü izdüşüm vektörü,

$$(\mathbf{y} \cdot \mathbf{U}_3) \mathbf{U}_3 = \left(\begin{bmatrix} 7 \\ 14 \\ 9 \end{bmatrix} \cdot \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \right) \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \\ -1 \end{bmatrix} \quad (4.5)$$

dır. Yani ortogonal ayrışımı, Şekil 4.2'de gösterildiği gibi;

$$\mathbf{y} = (\mathbf{y} \cdot \mathbf{U}_1) \mathbf{U}_1 + (\mathbf{y} \cdot \mathbf{U}_2) \mathbf{U}_2 + (\mathbf{y} \cdot \mathbf{U}_3) \mathbf{U}_3 \quad (4.6)$$

$$\begin{bmatrix} 7 \\ 14 \\ 9 \end{bmatrix} = 10 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -3.5 \\ 3.5 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0.5 \\ -1.0 \end{bmatrix} \quad (4.7)$$

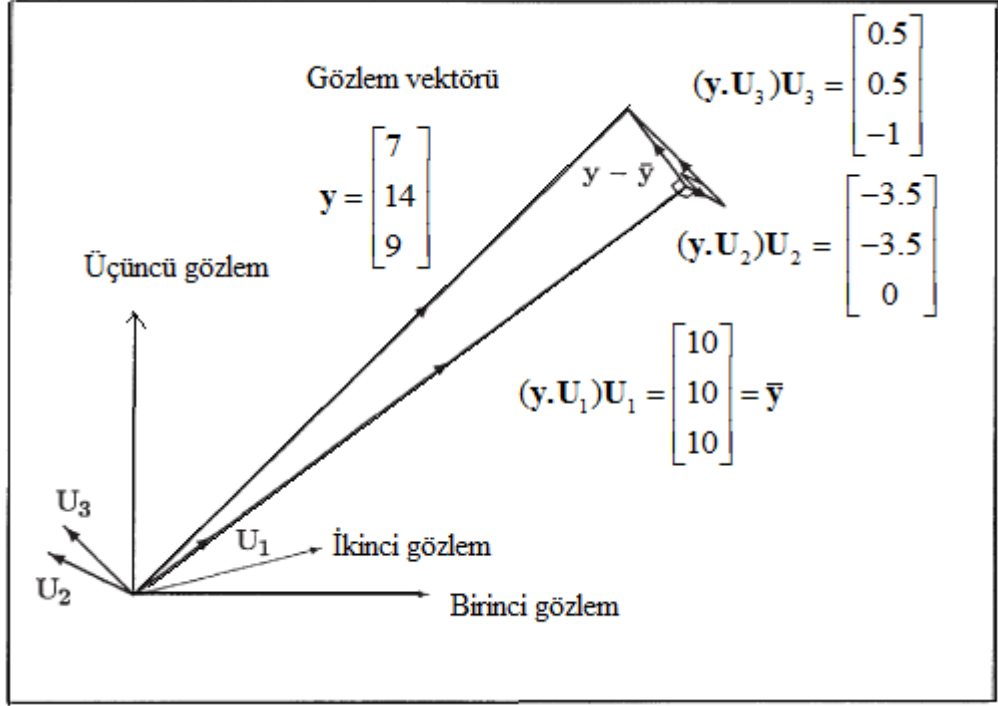
$\bar{y} = 10$ kalp atım hızındaki farkın en iyi tahmini olmak üzere, bu;

$$\begin{bmatrix} 7 \\ 14 \\ 9 \end{bmatrix} = 10 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -3.0 \\ 4.0 \\ -1.0 \end{bmatrix} \quad (4.8)$$

$$\mathbf{y} = \bar{\mathbf{y}} + (\mathbf{y} - \bar{\mathbf{y}}) \quad (4.9)$$

yani, gözlem vektörü = model vektörü + hata (tahmin edilen hata vektörü),

uygun modeline sadeleşir. Bu ayrışım aynı zamanda Şekil 4.2'de gösterilmiştir (Gözlemler ve örneklem ortalaması arasındaki farkları ifade etmek için $(\mathbf{y} - \bar{\mathbf{y}})$ 'yi kullandık).



Şekil 4.2. $(\mathbf{y} - \bar{\mathbf{y}})$ cinsinden gözlem vektörünün ortogonal ayrışımı

4.3. İzdüşüm Uzunluklarının Dağılımları

Kalp atım hızı ile ilgili gerçek ortalama farkının makul bir biçimde sıfır olup olmadığını test etmeye hazırlanmak için, $\mathbf{y} \cdot \mathbf{U}_1$, $\mathbf{y} \cdot \mathbf{U}_2$ ve $\mathbf{y} \cdot \mathbf{U}_3$ izdüşüm uzunluklarının dağılımlarını bilmemiz gerekir. Şimdi, y_1 , y_2 ve y_3 ; μ ortalamalı σ^2 varyanslı bir normal dağılımdan bağımsız gözlemler olmak üzere;

$$\mathbf{y} \cdot \mathbf{U}_1 = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \cdot \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{y_1 + y_2 + y_3}{\sqrt{3}} \quad (4.10)$$

dir.

Bölüm 3'deki gibi ilerleyerek, $\mathbf{y} \cdot \mathbf{U}_1$ ortalamasının $\sqrt{3}\mu$ ve $\mathbf{y} \cdot \mathbf{U}_1$ 'in varyansının σ^2 olduğunu buluruz. Yani $\mathbf{y} \cdot \mathbf{U}_1$; bir $N(\sqrt{3}\mu, \sigma^2)$ dağılımından gelir. Benzer şekilde, $\mathbf{y} \cdot \mathbf{U}_2$ ve $\mathbf{y} \cdot \mathbf{U}_3$ 'ün her ikisi de bir $N(0, \sigma^2)$ dağılımından gelir.

Bölüm 3'teki gibi, dikkat edilmesi gereken nokta, $\mathbf{y} \cdot \mathbf{U}_2$ ve $\mathbf{y} \cdot \mathbf{U}_3$ 'ün dağılımları daima sıfır noktası üzerinde merkezileşirken, $\mathbf{y} \cdot \mathbf{U}_1$ 'in dağılımı imkan dahilinde (potansiyel olarak) sıfırdan farklı bir nicelik (sayı) üzerinde merkezileşmiş olmasıdır.

4.4. Hipotezi Test Etme

Şimdi hedef çalışmayı inceleyeceğiz. Koşma kalp atışına etki eder mi? Bu konunun ön bilgisi olmadığını kabul ederek, 2-kuyruklu(2-yanlı) bir testi, yani, $H_1 : \mu \neq 0$ hipotezine karşı $H_0 : \mu = 0$ hipotezinin testini yapacağız. Burada μ benzer arka plana sahip şahısların sonsuz büyük bir teorik kitlesi için kalp atışındaki ortalama değişmedir.

Hipotez testimiz için, $(\mathbf{y} \cdot \mathbf{U}_1)^2$ izdüşüm karesinin, $\left[(\mathbf{y} \cdot \mathbf{U}_2)^2 + (\mathbf{y} \cdot \mathbf{U}_3)^2 \right] / 2$ hata doğrultuları için izdüşüm karelerinin ortalamasından önemli derecede daha büyük veya aynı olup olmadığını kontrol ederiz. Burada \mathbf{U}_2 ve \mathbf{U}_3 hata uzayını geren koordinat eksenleridir. İlgili Pisagor parçalanması, Şekil 4.3'te gösterildiği gibi,

$$\|\mathbf{y}\|^2 = (\mathbf{y} \cdot \mathbf{U}_1)^2 + (\mathbf{y} \cdot \mathbf{U}_2)^2 + (\mathbf{y} \cdot \mathbf{U}_3)^2 \quad (4.11)$$

veya

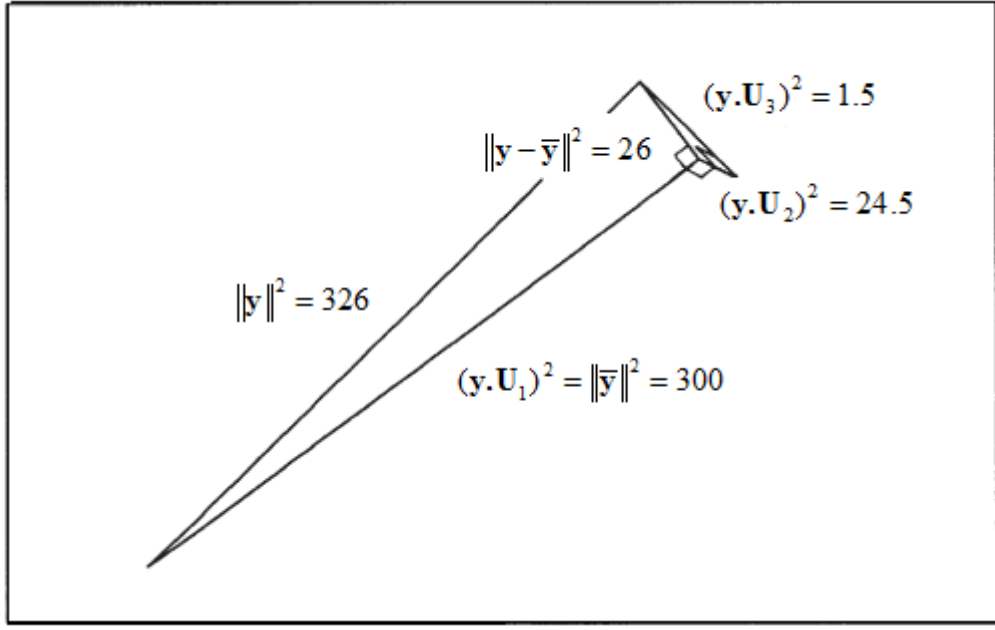
$$\|\mathbf{y}\|^2 = (\mathbf{y} \cdot \mathbf{U}_1)^2 + (\mathbf{y} \cdot \mathbf{U}_2)^2 + (\mathbf{y} \cdot \mathbf{U}_3)^2 \quad (4.12)$$

$$326 = 300 + 24.5 + 1.5$$

dir. Elde edilen test istatistiği,

$$F = \frac{(\mathbf{y} \cdot \mathbf{U}_1)^2}{\left[(\mathbf{y} \cdot \mathbf{U}_2)^2 + (\mathbf{y} \cdot \mathbf{U}_3)^2 \right] / 2} = 23.07 \quad (4.13)$$

dir.



Şekil 4.3. Üç ortogonal izdüşümün uzunluklarının karelerinin, Pisagor parçalanması

Bu gözlenen F değeri büyük veya küçük müdür? Şimdi eğer $\mu = 0$ ise $\mathbf{y} \cdot \mathbf{U}_1$, $\mathbf{y} \cdot \mathbf{U}_2$ ve $\mathbf{y} \cdot \mathbf{U}_3$ izdüşüm uzunluklarının tümü bir $N(0, \sigma^2)$ dağılımından gelir, bu nedenle, $(\mathbf{y} \cdot \mathbf{U}_1)^2 / \left\{ \left[(\mathbf{y} \cdot \mathbf{U}_2)^2 + (\mathbf{y} \cdot \mathbf{U}_3)^2 \right] / 2 \right\}$ test istatistiği, Ek 1'de tanımlanan $F_{1,2}$ dağılımından gelir. Bundan dolayı, 23.07'lik gözlenen F değerimizin büyük veya küçük olup olmadığını görmek için, onu $F_{1,2}$ dağılımının 90, 95 ve 99 yüzdeleri, yani, sırasıyla, 8.5, 18.5 ve 98.5 değerleriyle karşılaştırırız. 8.5 değerli 90 yüzdeliğinden ve 18.5 değerli 95 yüzdeliğinden daha büyük olduğundan, sonuç 23.07'lik gözlenen değer bize hipotezimiz ile ilgili bir fikir verir. Sıfır hipotezimizi $\alpha = 0.10$ ve $\alpha = 0.05$ düzeylerinde reddettiğimiz sonucuna varırız. Yani, veri incelemeye aldığımız kitledeki dinlenmiş ve koşan insanlar arasında, ortalama olarak kalp atım hızı ile ilgili bir fark olduğu sonucunu ortaya koyar.

4.5. İkili Örneklemeler t -Testi

$$F = (\mathbf{y} \cdot \mathbf{U}_1)^2 / \left\{ \left[(\mathbf{y} \cdot \mathbf{U}_2)^2 + (\mathbf{y} \cdot \mathbf{U}_3)^2 \right] / 2 \right\} = A^2 / [B^2 / 2] = 23.07 \quad (4.14)$$

test istatistiğimizin

$$t = \frac{\mathbf{y} \cdot \mathbf{U}_1}{\sqrt{[(\mathbf{y} \cdot \mathbf{U}_2)^2 + (\mathbf{y} \cdot \mathbf{U}_3)^2] / 2}} = \frac{A}{B / \sqrt{2}} = 4.80 \quad (4.15)$$

dönüştürülebildiğine dikkat ediniz. Önceden ifade edildiği gibi, eğer $\mu = 0$ ise, $\mathbf{y} \cdot \mathbf{U}_1$, $\mathbf{y} \cdot \mathbf{U}_2$ ve $\mathbf{y} \cdot \mathbf{U}_3$ izdüşüm uzunluklarının tümü bir $N(0, \sigma^2)$ dağılımından gelir.

Bu nedenle, $t = \mathbf{y} \cdot \mathbf{U}_1 / \sqrt{[(\mathbf{y} \cdot \mathbf{U}_2)^2 + (\mathbf{y} \cdot \mathbf{U}_3)^2] / 2}$ Ek 1'de tanımlanan t_2 - dağılımından, yani 2 serbestlik dereceli t - dağılımından, gelir. 4.80'lik gözlenen t değerimizin büyük veya küçük olup olmadığını görmek için, onu Ek 1'de verilen T tablosundaki t dağılımının yüzdeleri ile karşılaştırırız. Bu iki-yanlı dağılımın 95, 97.5 ve 99.5 yüzdeleri sırasıyla, 2.920, 4.303 ve 9.925'tir, bu nedenle, o; 95 yüzdeliği (%10 kritik değeri) aştığından, $\alpha = 0.10$ ve $\alpha = 0.05$ anlam düzeylerinde sıfır hipotezini reddederiz. Bundan dolayı, yine kitle ortalamasının sıfır olduğu düşüncesini reddetmek için bir delil bulduğumuz sonucunu çıkartırız.

4.6. Yorumlar

Yukarıdaki analizde \mathbf{U}_2 ve \mathbf{U}_3 doğrultuları σ^2 kitle varyansını tahmin etmek için kullanılırken, \mathbf{U}_1 doğrultusu μ kitle ortalamasını tahmin etmek için kullanılır. Uzun vade de, $\mathbf{y} \cdot \mathbf{U}_1$ izdüşüm uzunluğu $\sqrt{3}\mu$ ortalamasını alır, bu nedenle, $\bar{y} = 10; \mu$ 'yü tahmin etmemize yarar. Aynı zamanda uzun vade de, $(\mathbf{y} \cdot \mathbf{U}_2)^2$ ve $(\mathbf{y} \cdot \mathbf{U}_3)^2$ izdüşüm uzunlukları karesinin her biri σ^2 ortalamasını alır. Bu nedenle onların ortalaması, yani, $S^2 = 13; \sigma^2$ 'nin en iyi tahmin olarak işimize yarar. Ayrıca, uzun vadede, $(\mathbf{y} \cdot \mathbf{U}_1)^2$ izdüşüm uzunluğu karesinin $3\mu^2 + \sigma^2$ ortalamasını almaya çalıştığı gösterilebilir. Bu nedenle, eğer $\mu \neq 0$ ise $(\mathbf{y} \cdot \mathbf{U}_1)^2$ niceliği(sayısı) artırılmıştır ve $(\mathbf{y} \cdot \mathbf{U}_1)^2$, σ^2 'nin bir yansız tahmini ile bölünen artırılmış bir tahmini gibi düşünülebilir.

5. t- TESTİNE GEOMETRİK BİR BAKIŞ

Bir ikili örneklem veri kümesi için, Saville ve Wood (1996)'dan, yetişkin insanların üç karma cins ikiz çiftinde erkeklerin(M) ve kadınların(F) boylarını kullanırız. Birinci ikiz çiftte, Janet'in boyu 166 cm iken, John'un boyu 185 cm' idi. (Veri gerçektir fakat tüm isimler sahtedir.) İkinci ikiz çiftte Joanna'nın boyu 177.8 cm iken, Alistair'in boyu 185.4 cm idi. Üçüncü ikiz çiftte, Mary'nin boyu 160 cm iken, Bill'in boyu 182.9 cm idi. (M-F) boyundaki üç fark 19.0, 7.6 ve 22.9 cm'dir. Bu üç farkı, yetişkin insanların karma cins ikiz çiftlerindeki erkekler ve kadınlar arasındaki boyla ilgili farklardan oluşan, $N(\mu, \sigma^2)$ dağılımlı bir tek kitleden çekilen üç boyutlu örneklem olarak işleme sokarız.

Çizelge 5.1. Kadın ve erkeklerden oluşan 3 örneklem

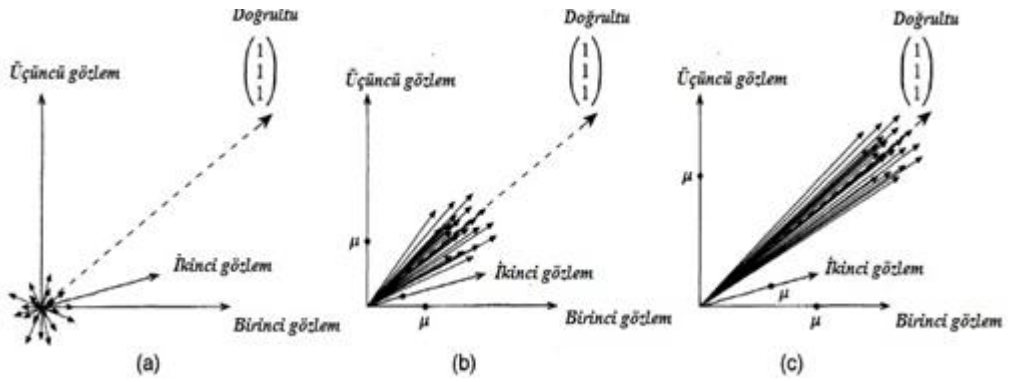
	1.çift	2.çift	3.çift
Erkek (M)	185	185.4	182.9
Kadın (F)	166	177.8	160
Fark (M-F)	19.0	7.6	22.9

5.1. Geometrik Yol

Yukarıdaki verinin geometrik yolu kullanan bir analizi için, hemen $H_0 : \mu = 0$ iki-yanlı test hipotezi altında yukarıdaki (19.0, 7.6, 22.9) örneği kadar aşırı veya daha aşırı bir örneklem elde etme şansını(olasılığını) nasıl elde ederiz? sorusunu sorarız. Fisher, tarafından keşfedilen çözüm yolu, veriyi üç boyutlu uzayda bir vektör, yani $(19.0, 7.6, 22.9)'$ vektörü olarak düşünmektir. Bu, doğal olarak aşırılığın bir ölçüsüne götürür.

Üç-boyutlu uzayda böyle veri vektörlerinin nasıl davrandığını anlamak için her bir şekilde bir incelemenin birçok tekrarından ortaya çıkan veri vektörlerini sergilediğimiz, Şekil 5.1'i göz önüne alınız. İlk olarak, eğer $\mu = 0$ ise, her bir

örneklem bir ilişkisiz üç değişkenli normal dağılımdan gelir. Bir dağılımın küresel simetrisi 3-boyutlu uzayda tüm doğrultuların büyük olasılıkla eş ölçüde (eş açılı) olduklarını garanti eder (Şekil 5.1(a)). İkinci olarak eğer $\mu \neq 0$ ise, veri vektörlerinin uçları $(\mu, \mu, \mu)'$ vektörünün ucu etrafında dağılırlar, bu nedenle, veri vektörlerinin doğrultuları $(1,1,1)'$ doğrultusuna daha yakındırlar (Şekil 5.1(b) ve Şekil 5.1(c)). Son olarak, daha büyük μ, σ ile ilgilidir, yani her bir veri vektörü ve $(1,1,1)'$ doğrultusu arasındaki açı daha küçüktür (Şekil 5.1(b)'ye karşı Şekil 5.1(c)).



Şekil 5.1. (a) $\mu = 0$, (b) $\mu \neq 0$ ve (c) $\mu \neq 0$ durumlarında bir incelemenin birçok tekrarından ortaya çıkan veri vektörleri¹

Bu, $(19.0, 7.6, 22.9)'$ özel veri vektörümüz ve $(1,1,1)'$ doğrultusu arasındaki θ açısını, örneğimizin ne kadarının $H_0 : \mu = 0$ hipotezi altında olduğunun bir ölçüsü olarak kullanabileceğimizi belirtir. Eğer açı “büyük” ise, bu takdirde verilerimiz $H_0 : \mu = 0$ ile tutarlıdır, eğer açı “küçük” ise, bu takdirde verilerimiz $H_1 : \mu \neq 0$ karşıt (alternatif) hipotezi ile tutarlıdır.

¹ (a) $\mu = 0$ (b) $\mu \neq 0$ ve μ (b) dekinden daha büyük olmak üzere (c), $\mu \neq 0$ durumlarında incelemenin birçok tekrarından sonuçlanan veri vektörleri: Genel olarak, bir veri vektörü ve $(1,1,1)'$ doğrultusu arasındaki açı, μ daha büyük iken, daha küçük olur; bu, $H_0 : \mu = 0$ hipotezi için açının bir test istatistiği olarak kullanımını sezgi yoluyla ortaya koyar.

Özel örneğimiz için, iki vektörün nokta çarpımını (iç-çarpımını) içeren, lineer cebirden bir standart formülü kullanarak θ 'yı aşağıdaki gibi hesaplarız.

$$\cos \theta = \frac{\begin{pmatrix} 19.0 \\ 7.6 \\ 22.9 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}{\sqrt{(19.0)^2 + (7.6)^2 + (22.9)^2} \sqrt{3}} = \frac{19.0 + 7.6 + 22.9}{\sqrt{2829.51}} = 0.93057 \quad (5.1)$$

Bu nedenle, açı Şekil 5.2'de gösterildiği gibi, $\theta = 21.476^\circ$ (veya 0.3748 radyan) dir. Bu açı büyük veya küçük müdür? Bunu cevaplamak için $H_0 : \mu = 0$ hipotezi altında 21.476° kadar veya ondan daha küçük θ açısı ile bir veri vektörünü gözlemlemenin p olasılığını hesaplarız. Bu, veri vektörünü $(1,1,1)$ ' doğrultusu etrafında döndürmek suretiyle oluşturulan (sonsuz) double (çifte) koni içinde uzanmasının olasılığıdır (bkz. Şekil 5.2). p değerini tayin etmek için, ilk olarak, H_0 hipotezi altında $r_0 = \sqrt{(19.0)^2 + (7.6)^2 + (22.9)^2} = 30.711$ gibi sabit uzunluklu bir veri vektörünün, (sonsuz) çifte koni içerisinde uzanması olasılığını hesaplarız. H_0 hipotezi altında, veri vektör doğrultularının dağılımının küresel simetrisinden, bu olasılık,

$$\frac{\text{yüzey alanı(şekil 2 deki taralı kesişim)}}{\text{yüzey alanı}(r_0 \text{ yarıçaplı küre})} = \frac{4\pi r_0^2 \{1 - \cos(\theta)\}}{4\pi r_0^2} = 1 - \cos(\theta) \quad (5.2)$$

dır. Burada yüzey alanları basit analizle elde edildi ve cevap hesaplamada kullanılan r_0 - değerinden bağımsızdır.

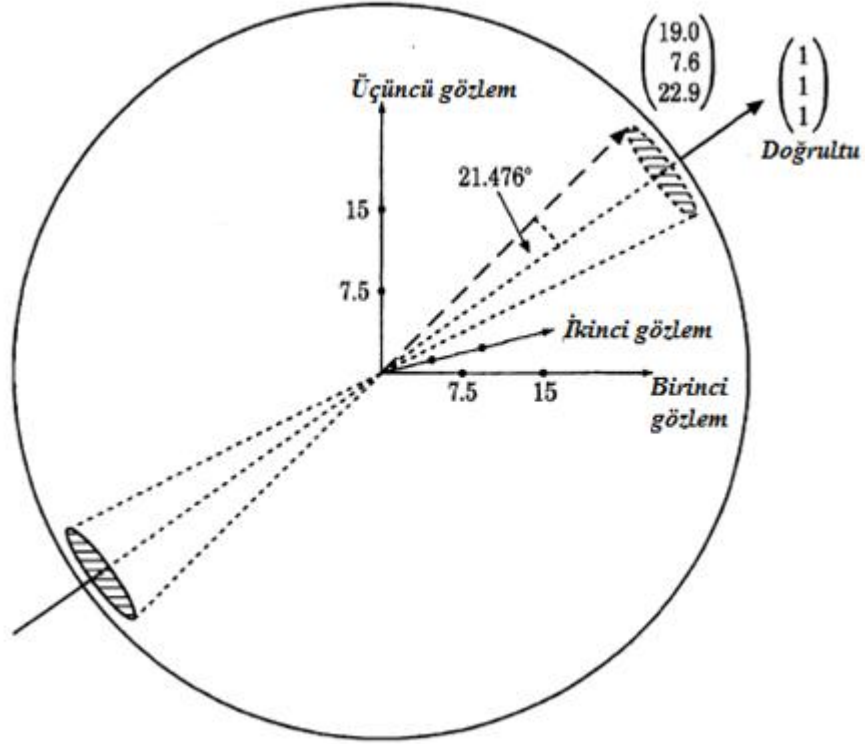
p-değerinin hesaplanmasını tamamlamak için şimdi yukarıdaki şartlı olasılığı bir veri vektörünün r uzunluğa sahip olması olasılığı ile ağırlıklandırarak, veri vektörünün mümkün olabilir tüm r uzunlukları üzerinden integral alırız. Bu,

$$\begin{aligned} p &= \int_0^\infty \{1 - \cos(\theta)\} f(r) dr = \{1 - \cos(\theta)\} \int_0^\infty f(r) dr = 1 - \cos(\theta) \\ &= 1 - \cos(21.476) = 0.07 \end{aligned} \quad (5.3)$$

olduğunu ortaya koyar. Burada $f(r)$, bir veri vektörünün uzunluğu için olasılık yoğunluk fonksiyonudur. Analizimizi sonuçlandırmak için 0.07'lik p-değerinin

0.05'ten daha büyük olduğuna, bu nedenle, eğer 0.05 kesim noktasını kriterimiz (kritik nokta) olarak kullanırsak, $H_0 : \mu = 0$ hipotezi altında verimizin olağandışı olmadığını farkına varırız.

Karma cins ikiz çiftlerde erkek ve kadın arasında boyca sıfırdan farklı bir gerçek ortalama farkının güçlü delilini elde edemediğimiz kararına varırız.



Şekil 5.2. $r_0 = \sqrt{(19.0)^2 + (7.6)^2 + (22.9)^2} = 30.711$ yarıçaplı küreye ilişkin $(19.0, 7.6, 22.9)'$ veri vektörü²

² $r_0 = \sqrt{(19.0)^2 + (7.6)^2 + (22.9)^2} = 30.711$ yarıçaplı küreye ilişkin $(19.0, 7.6, 22.9)'$ veri vektörü ve küre ile kesişimi taralı bölge olan çifte koni taralıdır: Taralı kesişimin yüzey alanının kürenin yüzey alanına oranı, μ 'nın sıfırdan farklı olduğunun güçlü delilidir.

5.2. Geometrik ve Geleneksel Yollar Arasındaki Bağlantı

Yukarıdaki alt kısımda t -değerinden söz etmeksizin, $H_0 : \mu = 0$ hipotezinin ikili örneklem t -testini tamamladık. p değeri için geometrik yol daha geleneksel yola nasıl bağlanır?

Şimdi ikili örneklem için bilinen t -test istatistiği;

$$\begin{aligned} t &= \frac{\bar{y}}{s / \sqrt{n}} = \frac{\bar{y}\sqrt{n}}{\sqrt{\left\{ \sum_{i=1}^n (y_i - \bar{y})^2 / (n-1) \right\}}} \\ &= \frac{16.5\sqrt{3}}{\sqrt{\left[\{2.5^2 + (-8.9)^2 + 6.4^2\} / 2 \right]}} = 3.595 \end{aligned} \quad (5.4)$$

dir. Burada,

y_1, y_2, \dots, y_n : veri değerleri (İkili örneklem değerleri arasındaki farklar)

\bar{y} : örneklem ortalaması

s : örneklem standart yapması

n : örneklem büyüklüğünü ifade eder.

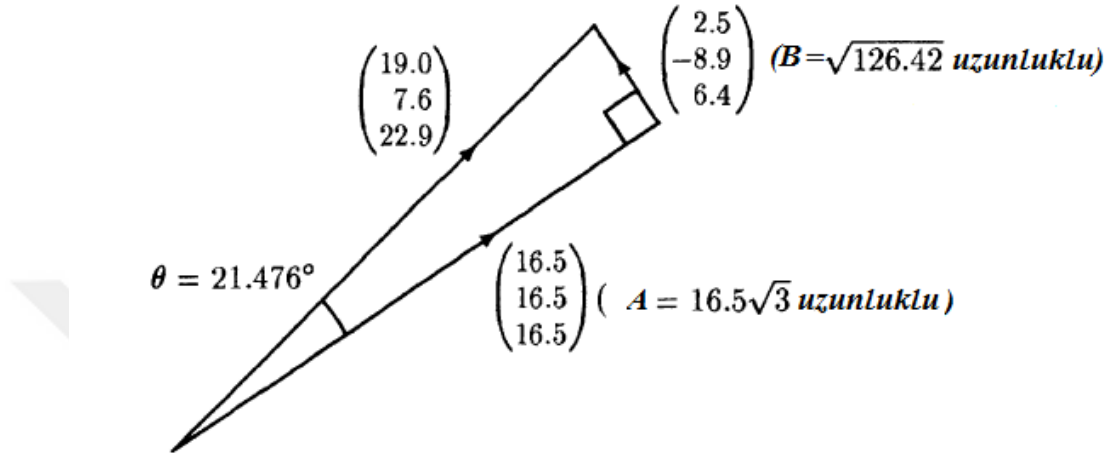
Bunu θ ya bağlamak için, Şekil 5.2'de dolaylı olarak anlatılan bir dik üçgeni Şekil 5.3'te gösteririz. Bu üçgen veri vektörünü $(1,1,1)'$ doğrultusu üzerine izdüşürmek suretiyle elde edilir ve $A = \sqrt{3}(16.5)^2 = 16.5\sqrt{3}$ ve $B = \sqrt{\{2.5^2 + (-8.9)^2 + 6.4^2\}} = \sqrt{126.42}$ uzunluğuna sahip kenarlar ile

$$\begin{pmatrix} 19.0 \\ 7.6 \\ 22.9 \end{pmatrix} = \begin{pmatrix} 16.5 \\ 16.5 \\ 16.5 \end{pmatrix} + \begin{pmatrix} 2.5 \\ -8.9 \\ 6.4 \end{pmatrix} \quad (5.5)$$

vektör toplamını gösterir. Bu nedenle üç büyüklüğüne sahip bir örneklemin durumu için t ve θ arasındaki $t = \sqrt{2} \cot(\theta)$ ilişkisini saptayan

$$t = \frac{16.5\sqrt{3}}{\sqrt{[(2.5)^2 + (-8.9)^2 + 6.4^2] / 2}} = \frac{A}{B/\sqrt{2}} = \sqrt{2} \cot(\theta) = \sqrt{2} \cot(21.476) = 3.595 \quad (5.6)$$

dir. Bu arada örneklem ortalaması ve standart sapmasının $A = \bar{y}\sqrt{3}$ ve $B = s\sqrt{2}$ olmak üzere üçgen içinde kendiliğinden ortaya çıktığına dikkat ediniz.



Şekil 5.3. Veri vektörünün vektör ayrışımı³

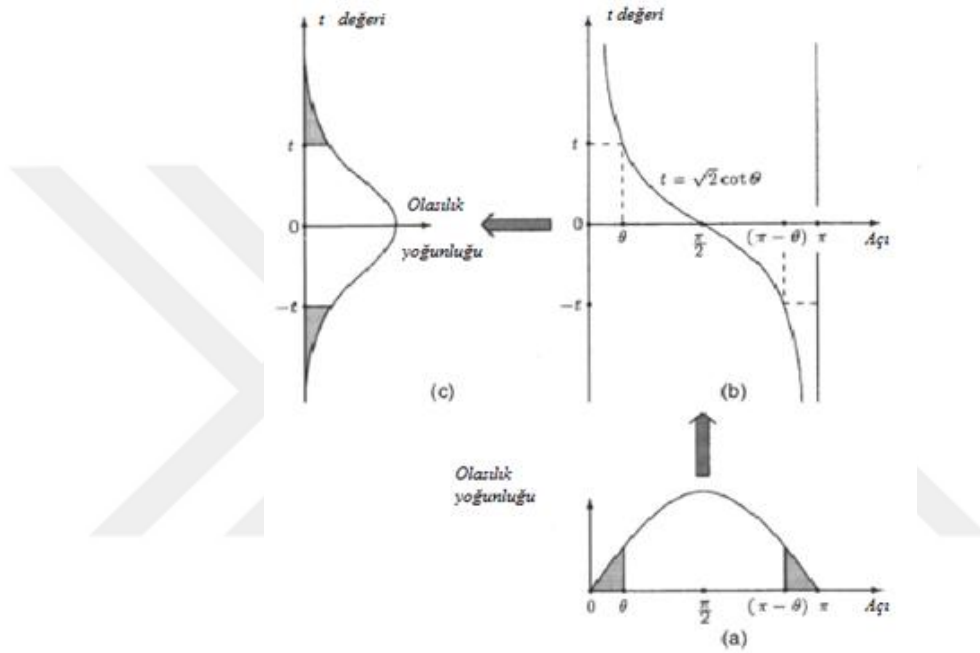
5.3. θ 'dan t-ye Dönüşüm

Yukarıdaki gelişimde açıklamanın kolaylığı için, θ açısı tam olarak $0 - 90^\circ$ değişim aralığına kısıtlandı. Bu iki-yanlı testler için tam anlamıyla uygun değildir, fakat bir-yanlı testler için uygun değildir. Daha genel olarak, θ , veri vektörü ve $(1,1,1)'$ doğrultusunun pozitif yönü arasındaki açı olarak tanımlanır, bu nedenle, $0^\circ \leq \theta \leq 180^\circ$ 'dir. Eğer θ 'yı radyan ($0 \leq \theta \leq \pi$) olarak da ifade edersek, bu takdirde

$$\int_0^\theta \frac{\sin(u)}{2} du = \frac{1 - \cos(\theta)}{2} \quad (5.7)$$

³Veri vektörünün vektör ayrışımı: Bilinen t- istatistiği $\sqrt{2} \cot(\theta)$ 'ya eşittir.

olduğundan, basit biçimiyle θ için olasılık yoğunluk fonksiyonunu $\sin(\theta)/2$ olarak yazabiliriz. Kısaca $H_0 : \mu = 0$ (şekil 5.4(a)) hipotezi altında θ 'nın(radyan olarak) dağılımı $t = \sqrt{2} \cot(\theta)$ dönüşümü ile (bkz. şekil 5.4(b)) t_2 dağılımına(bkz. şekil 5.4 (c)) çevrilir. θ 'nın sıfır veya π ye yakın değerlerinin sırasıyla t 'nin büyük pozitif ve negatif değerlerine dönüşürken, θ 'nın $\pi/2$ 'ye yakın değerlerinin t 'nin küçük değerlerine dönüştüklerine dikkat ediniz.



Şekil 5.4. $t = \sqrt{2} \cot(\theta)$ dönüşümü⁴

5.4. $H_0 : \mu = \mu_0$ Durumu

$y - \mu_0$ dönüştürülmüş değişkenini kullanarak yürütülen veri analizi ile, $H_0 : \mu = \mu_0$ ($\neq 0$) daha genel durumunun, şimdiki durumda $H_0 : \mu - \mu_0 = 0$ olarak yeniden yazılabildiğine dikkat ederiz.

⁴ θ için olasılık yoğunluk fonksiyonu sırasıyla (a) ve (c) de gösterilen t için olasılık yoğunluk fonksiyonuna bağlayan $t = \sqrt{2} \cot(\theta)$ dönüşümü: Taralı olanlar θ veya t 'den daha uçta olan değerleri gösterir.

BÖLÜM 6. BAĞIMSIZ ÖRNEKLEMLER

Gözlemlerin doğal bir eşlemesi olmadığında bağımsız örneklemeler ortaya çıkar. Örneğin, altı koşucu ve altı dinlenmiş insanın kalp atışları ölçüsü rasgele seçilebilir. İlgimiz normal olarak, dingin(dinlenmiş) ve koşan insanlar arasında kalp atışı ile ilgili ortalama farktır. Daha biçimsel olarak, iki inceleme kitlemizin μ_1 ve μ_2 ortalamalarını karşılaştırmayla ilgileniriz. Analiz için, her iki kitlenin de normal dağıldıklarını ve iki kitlenin varyanslarının aynı olduklarını ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) kabul ederiz.

Aşağıdaki kısımda dingin ve koşan insanlar arasında kalp atışı ile ilgili bir veri kümesini tanımlayacağız. Bu veri kümesinin iki koşucu ve iki dingin insandan oluşan bir alt kümesi analiz edilecektir.

6.1. Veri Kümesi

Bölüm 4'te her bir kişi için koşudan önceki ve sonraki kalp atışındaki değişme kaydedildi. Bu bölümde inceleme amaçlı farklı bir tasarımı göz önüne alacağız. Yeni tasarım altında 26 büyüklüğüne sahip rasgele iki gruba bölünen 52 kişiye sahip olduğumuzu varsayalım. Birinci gruptaki şahıslar 10 dakika hareketsiz oturduktan sonra 30 saniye üzerinden kalp atışlarını kaydetti. İkinci gruptaki şahıslar olay yerinde 2 dakika koştuktan sonra hemen kalp atışlarını kaydetti (Saville ve Wood, 1996). Sonuçta elde edilen veriler aşağıdaki iki çizelge de verilmiştir.

Çizelge 6.1. Bölüm 3'teki 26 tarım araştırmacısının koşudan önce ve sonraki gerçek kalp atışları

a) Grup 1: Hareketsiz şahıslar

Şahıs	Kalp atım hızı	Şahıs	Kalp atım hızı
1	24	14	30
2	40	15	32
3	34	16	38
4	35	17	31
5	31	18	32
6	37	19	37
7	40	20	27
8	33	21	40
9	35	22	29
10	27	23	35
11	26	24	32
12	35	25	41
13	26	26	41

b) Grup 2: Koşu sonrası şahıslar

Şahıs	Kalp atım hızı	Şahıs	Kalp atım hızı
1	34	14	44
2	40	15	44
3	57	16	42
4	46	17	27
5	44	18	43
6	38	19	54
7	44	20	44
8	60	21	47
9	45	22	38
10	47	23	43
11	38	24	40
12	39	25	37
13	43	26	48

6.2. İki Büyüklüğüne Sahip Örneklemeler

Bu kısımda, geometrik yöntemimizi, iki kitlenin ortalamalarını karşılaştırmak istediğimiz duruma uyguluyoruz ve iki boyutlu iki bağımsız rasgele örneklem emrimizde olur. Örneğimiz için Çizelge 6.2’de gösterilen veriyi kullanacağız; bu veriler Çizelge 6.1’den iki dingin kişi ve iki koşu sonrası kişinin kalp atım hızlarıdır.

Çizelge 6.2. İki dingin ve iki koşu sonrası kişinin kalp atım hızları ile örneklem ortalaması

Kalp atım hızı(30 saniye)		Ortalama
Dingin kişi	31 37	34
Koşu sonrası kişi	44 47	45,5

Elde edilen gözlem vektörü

$$\mathbf{y} = \begin{bmatrix} 31 \\ 37 \\ 44 \\ 47 \end{bmatrix} \quad (6.1)$$

dir.

6.2.1. Amaç

İncelemenin amacı:“Koşma kalp atışını etkiler mi?” sorusudur. Bu soruyu cevaplamak için, biçimsel olarak $H_1 : \mu_1 \neq \mu_2$ hipotezine karşı $H_0 : \mu_1 = \mu_2$ hipotezini test ederiz.

6.2.2. Temel Düşünce

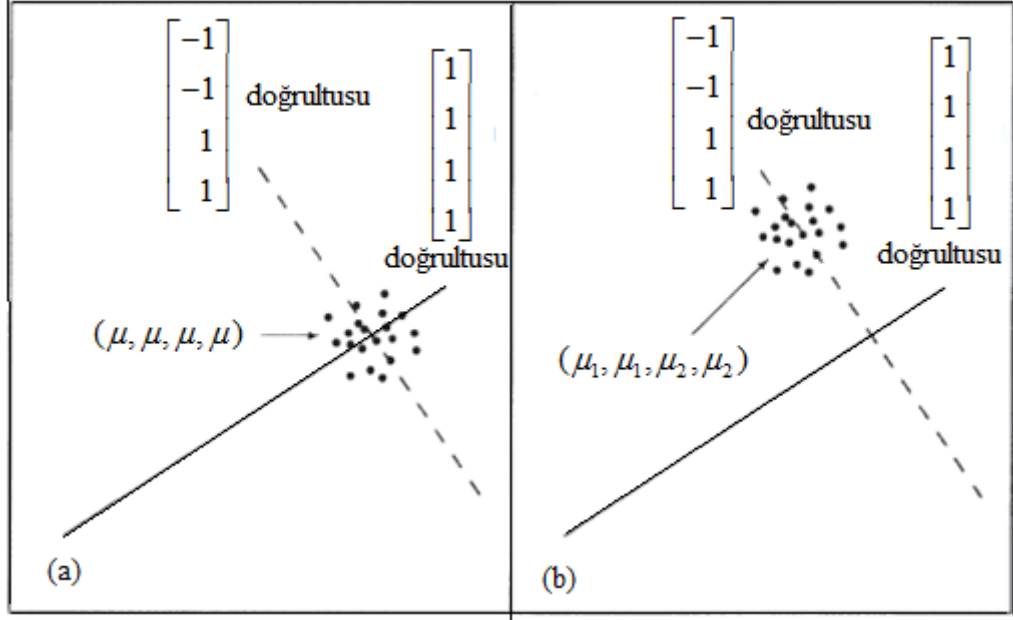
Şekil 6.1 yeni durumumuz için temel düşünceyi açıklar. Eğer kalp atım hızı bakımından, dingin ve koşu sonrası kişiler, ortalama olarak aynı iseler ($\mu_1 = \mu_2$), bu takdirde $\mu_1 = \mu_2 = \mu$ olmak üzere, gözlem vektörümüz Şekil 6.1(a)’da gösterildiği gibi (μ, μ, μ, μ) noktası üzerinde merkezlenen bir dağılımın parçası olacak. Bununla beraber, eğer onlar kalp atım hızı bakımından farklı iseler ($\mu_1 \neq \mu_2$), bu takdirde gözlem vektörümüz Şekil 6.1(b)’de gösterildiği gibi, $(\mu_1, \mu_1, \mu_2, \mu_2)$

noktası üzerinde merkezlenen bir dağılımın(saçılımın) parçası olacak. Başka bir deyişle, eğer $\mu_1 = \mu_2$ ise, bu takdirde dağılım eş açılı doğru üzerinde merkezlenir.

Halbuki, $\mu_1 \neq \mu_2$ ise, bu takdirde $\mu = \frac{(\mu_1 + \mu_2)}{2}$ olduğundan, dağılım, eş açılı doğrudan

$$\begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \end{bmatrix} - \begin{bmatrix} \mu \\ \mu \\ \mu \\ \mu \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu \\ \mu_1 - \mu \\ \mu_2 - \mu \\ \mu_2 - \mu \end{bmatrix} = \begin{bmatrix} (\mu_1 - \mu_2)/2 \\ (\mu_1 - \mu_2)/2 \\ (\mu_2 - \mu_1)/2 \\ (\mu_2 - \mu_1)/2 \end{bmatrix} = \frac{\mu_2 - \mu_1}{2} \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \quad (6.2)$$

doğrultusunda uzaklaşır.



Şekil 6.1. Kalp atım hızını temsil eden veri noktalarının bulutları

6.2.3. Uygun Bir Koordinat Sistemi

Amacımız için bu takdirde 4- boyutlu uzay için koordinat eksenlerinin uygun bir kümesi:

$$\mathbf{U}_1 = \frac{1}{\sqrt{4}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{U}_2 = \frac{1}{\sqrt{4}} \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{U}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{U}_4 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \quad (6.3)$$

dır. Burada $\mathbf{U}_2 = [-1, -1, 1, 1]' / \sqrt{4}$ ilgilenilen $(\mu_2 - \mu_1)$ farkı ile ilgili doğrultu iken, $\mathbf{U}_1 = [1, 1, 1, 1]' / \sqrt{4}$ genel ortalamayla, yani, μ ilgili olan doğrultudur.

6.2.4. Ortogonal Ayrışım

Şimdi, $\mathbf{y} = [31, 37, 44, 47]'$ gözlem vektörümüzü koordinat eksenlerimizin her biri üzerine izdüşürürüz. Birinci izdüşüm vektörü

$$(\mathbf{y} \cdot \mathbf{U}_1) \mathbf{U}_1 = \left(\begin{bmatrix} 31 \\ 37 \\ 44 \\ 47 \end{bmatrix} \cdot \frac{1}{\sqrt{4}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right) \frac{1}{\sqrt{4}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 39.75 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad (6.4)$$

ikinci izdüşüm vektörü

$$(\mathbf{y} \cdot \mathbf{U}_2) \mathbf{U}_2 = \left(\begin{bmatrix} 31 \\ 37 \\ 44 \\ 47 \end{bmatrix} \cdot \frac{1}{\sqrt{4}} \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \right) \frac{1}{\sqrt{4}} \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix} = 5.75 \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \quad (6.5)$$

üçüncü izdüşüm vektörü

$$(\mathbf{y} \cdot \mathbf{U}_3) \mathbf{U}_3 = \left(\begin{bmatrix} 31 \\ 37 \\ 44 \\ 47 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right) \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = 3 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad (6.6)$$

ve dördüncü izdüşüm vektörü

$$(\mathbf{y} \cdot \mathbf{U}_4) \mathbf{U}_4 = \left(\begin{bmatrix} 31 \\ 37 \\ 44 \\ 47 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right) \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} = 1.5 \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \quad (6.7)$$

biçimindedir. Gözlem vektörünün karşılık gelen ortogonal ayrışım

$$\mathbf{y} = (\mathbf{y} \cdot \mathbf{U}_1)\mathbf{U}_1 + (\mathbf{y} \cdot \mathbf{U}_2)\mathbf{U}_2 + (\mathbf{y} \cdot \mathbf{U}_3)\mathbf{U}_3 + (\mathbf{y} \cdot \mathbf{U}_4)\mathbf{U}_4$$

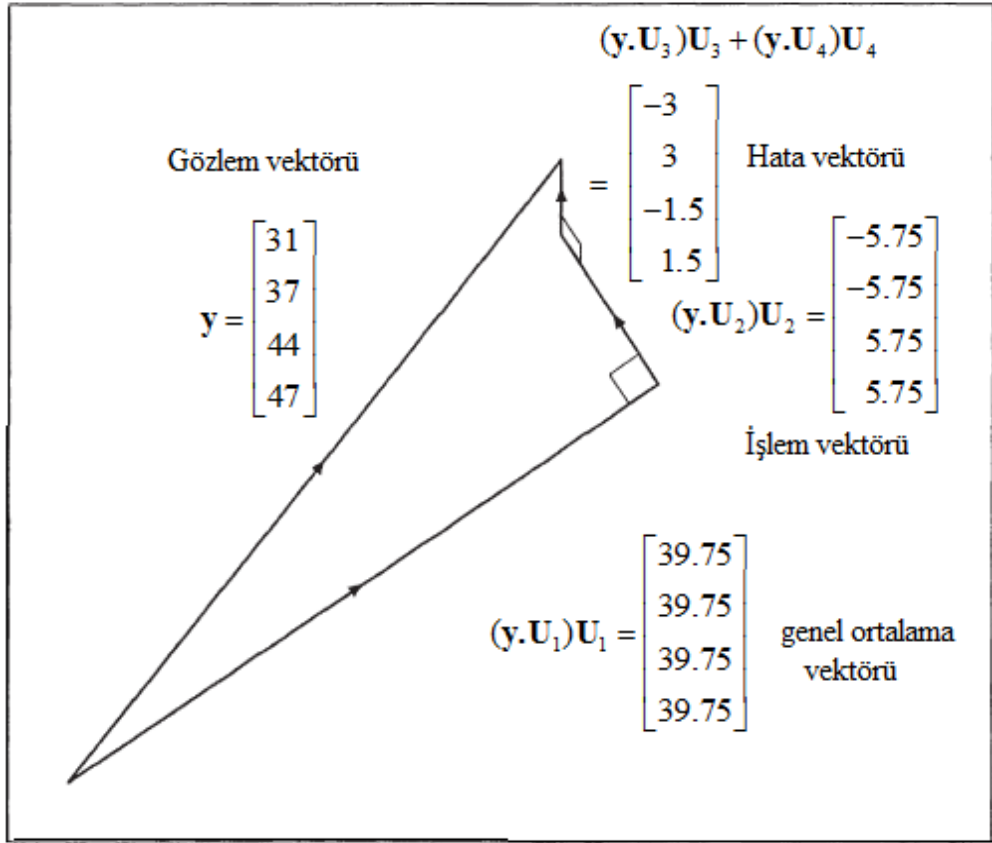
$$\begin{bmatrix} 31 \\ 37 \\ 44 \\ 47 \end{bmatrix} = (159/4)\mathbf{U}_1 + (23/4)\mathbf{U}_2 + (6/2)\mathbf{U}_3 + (3/2)\mathbf{U}_4$$

$$\begin{bmatrix} 31 \\ 37 \\ 44 \\ 47 \end{bmatrix} = \begin{bmatrix} 39.75 \\ 39.75 \\ 39.75 \\ 39.75 \end{bmatrix} + \begin{bmatrix} -5.75 \\ -5.75 \\ 5.75 \\ 5.75 \end{bmatrix} + \begin{bmatrix} -3 \\ 3 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -1.5 \\ 1.5 \end{bmatrix} \quad (6.8)$$

dır. Bu,

$$\begin{bmatrix} 31 \\ 37 \\ 44 \\ 47 \end{bmatrix} = \underbrace{\begin{bmatrix} 39.75 \\ 39.75 \\ 39.75 \\ 39.75 \end{bmatrix}}_{\text{Genel ortalama vektörü}} + \underbrace{\begin{bmatrix} -5.75 \\ -5.75 \\ 5.75 \\ 5.75 \end{bmatrix}}_{\text{İşlem vektörü}} + \underbrace{\begin{bmatrix} -3 \\ 3 \\ -1.5 \\ 1.5 \end{bmatrix}}_{\text{Hata vektörü}} \quad (6.9)$$

uygun modeline sadeleşir. Yani, \mathbf{y} vektörümüzü Şekil 6.2’te gösterildiği gibi, bir genel ortalama vektörü, bir işlem vektörü ve bir hata vektörü olmak üzere, üç ortogonal bileşenlere ayırdık. İşlem terimini burada oldukça uzun bir “incelemenin kitleleri arasındaki farklar” deyişi için amacımıza uygun iyi bir kısaltma olarak kullanırız.



Şekil 6.2. İki inceleme kitlesinden alınan iki büyüklüğüne sahip bağımsız örneklem durumu için gözlem vektörünün ortogonal ayrışımı

6.2.5. İzdüşüm Uzunluklarının Dağılımları

Kalp atım hızı ile ilgili gerçek ortalamanın sıfır olup olmadığını test etmeye hazırlanmak için, şimdi $y \cdot U_1$, $y \cdot U_2$, $y \cdot U_3$ ve $y \cdot U_4$ izdüşüm uzunluklarının dağılımlarını ortaya koyacağız. İlk olarak,

$$y \cdot U_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} \cdot \frac{1}{\sqrt{4}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{y_{11} + y_{12} + y_{21} + y_{22}}{\sqrt{4}} \quad (6.10)$$

elde ederiz. Burada y_{11} ve y_{12} sembollerini, birinci ve ikinci gözlemlerin birinci inceleme kitlesinden (μ ortalama ve σ^2 varyanslı) geldiğini ifade etmek için kullanırız. y_{21} ve y_{22} sembollerini birinci ve ikinci gözlemlerin ikinci inceleme kitlesinden (μ ortalama ve σ^2 varyanslı) geldiğini ifade etmek için kullanırız.

Bu nedenle, Şekil 6.3'te gösterildiği gibi,

$$\mathbf{y} \cdot \mathbf{U}_1 = \frac{\mu_1 + \mu_2 + \mu_3 + \mu_4}{\sqrt{4}} = \frac{4\mu}{\sqrt{4}} = 2\mu, \quad (6.11)$$

ikinci olarak,

$$\mathbf{y} \cdot \mathbf{U}_2 = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} \cdot \frac{1}{\sqrt{4}} \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix} = \frac{-y_{11} - y_{12} + y_{21} + y_{22}}{\sqrt{4}} \quad (6.12)$$

dür. Böylece $\mathbf{y} \cdot \mathbf{U}_2$ 'nin ortalaması

$$\mathbf{y} \cdot \mathbf{U}_2 = \frac{-\mu_1 - \mu_1 + \mu_2 + \mu_2}{\sqrt{4}} = \mu_2 - \mu_1 \quad (6.13)$$

dir. Üçüncü olarak,

$$\mathbf{y} \cdot \mathbf{U}_3 = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \frac{-y_{11} + y_{12}}{\sqrt{2}} \quad (6.14)$$

dır. Bu nedenle, $\mathbf{y} \cdot \mathbf{U}_3$ 'ün ortalaması;

$$\mathbf{y} \cdot \mathbf{U}_3 = \frac{-\mu_1 + \mu_1}{\sqrt{2}} = 0 \text{ 'dır.} \quad (6.15)$$

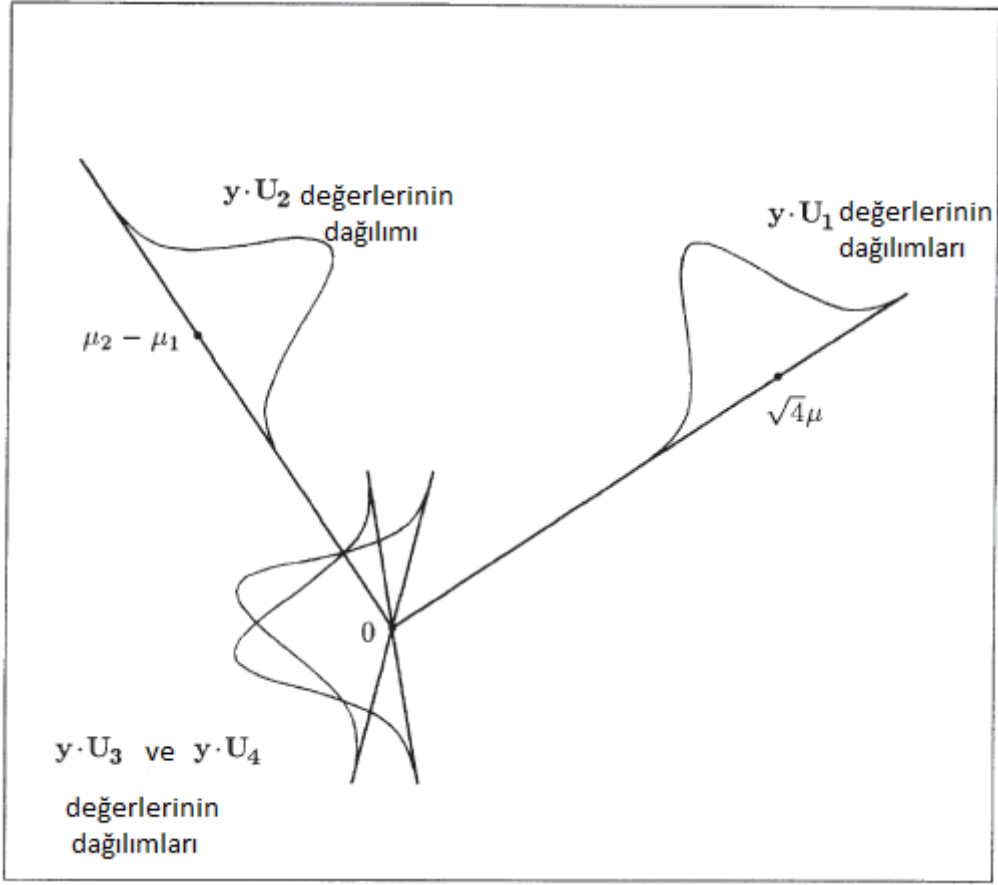
Son olarak,

$$\mathbf{y} \cdot \mathbf{U}_4 = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} = \frac{-y_{21} + y_{22}}{\sqrt{2}} \quad (6.16)$$

dır. Bu nedenle, $\mathbf{y} \cdot \mathbf{U}_4$ 'ün ortalaması;

$$\mathbf{y} \cdot \mathbf{U}_4 = \frac{-\mu_2 + \mu_2}{\sqrt{2}} = 0 \quad (6.17)$$

olarak ortaya konur.



Şekil 6.3. $y.U_1, y.U_2, y.U_3$ ve $y.U_4$ izdüşüm uzunluklarının dağılımlarının resimsel gösterimi

Bundan başka, dört izdüşüm uzunluğunun hepsinin σ^2 varyansına sahip oldukları kolayca gösterilebilir. Bu nedenle, Şekil 6.3'te gösterildiği gibi, $y.U_1$ bir $N(2\mu, \sigma^2)$ dağılımından, $y.U_2$ 'nin bir $N(\mu_1 - \mu_2, \sigma^2)$ dağılımından ve $y.U_3$ ile $y.U_4$ 'ün her ikisinin de bir $N(0, \sigma^2)$ dağılımından geldiği görülür. Dikkat edilmesi gereken husus: $y.U_3$ ve $y.U_4$ daima sıfır üzerinde merkezlenirken, $y.U_1$ ve $y.U_2$ 'nin potansiyel olarak sıfırdan farklı nicelikler üzerinde merkezlendikleridir.

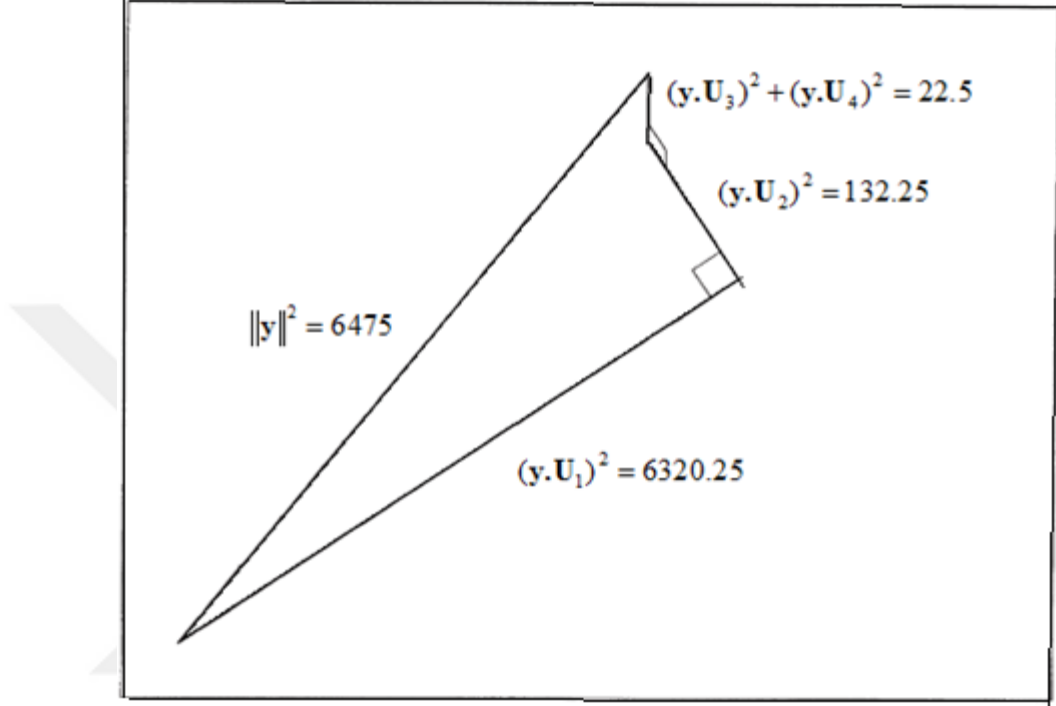
6.2.6. Hipotezi Test Etme

Şimdi, inceleme amacımızı araştırmayı sürdüreceğiz: Hareketsiz kişiler ile koşudan sonraki kişiler arasında kalp atım hızı ilgili gerçek ortalama fark sıfır mıdır? Daha biçimsel olarak, $H_1 : \mu_1 \neq \mu_2$ alternatif hipotezine karşı, $H_0 : \mu_1 = \mu_2$ sıfır hipotezini test etmeyi istiyoruz. Uygun Pisagor parçalanması Şekil 6.4'te gösterildiği gibi,

$$\|y\|^2 = (y \cdot U_1)^2 + (y \cdot U_2)^2 + (y \cdot U_3)^2 + (y \cdot U_4)^2 \quad (6.18)$$

$$6475 = 6320.25 + 132.25 + 18 + 4.5$$

dır.



Şekil 6.4. Gözlem vektörünün uzunluğunun karesinin Pisagor parçalanması⁵

Hipotezi test etmek için, sadece $(y \cdot U_2)^2$ 'nin izdüşüm uzunluğu karesinin, $(y \cdot U_3)^2$ ve $(y \cdot U_4)^2$ izdüşüm uzunlukları karelerinin ortalaması ile aynı veya ondan önemli derecede daha büyük olup olmadığını kontrol ederiz.

⁵Bu diyagram ölçekli değildir.

Elde edilen test istatistiği ;

$$F = \frac{(\mathbf{y} \cdot \mathbf{U}_2)^2}{\left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + (\mathbf{y} \cdot \mathbf{U}_4)^2 \right] / 2} = \frac{132.25}{[18+4.5]/2} = \frac{132.25}{11.25} = 11.75$$

dır.

Bu gözlenen F değeri büyük ya da küçük müdür? Şimdi eğer $\mu_1 = \mu_2$ ise, $\mathbf{y} \cdot \mathbf{U}_2$, $\mathbf{y} \cdot \mathbf{U}_3$ ve $\mathbf{y} \cdot \mathbf{U}_4$ izdüşüm uzunluklarının hepsi bir $N(0, \sigma^2)$ dağılımından gelir, bu nedenle, $F = (\mathbf{y} \cdot \mathbf{U}_2)^2 / \left\{ \left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + (\mathbf{y} \cdot \mathbf{U}_4)^2 \right] / 2 \right\}$ test istatistiği, Ek 1'de tanımlanan $F_{1,2}$ dağılımından gelir. Bundan dolayı, 11.75'lik gözlenen değerimizin büyük veya küçük olup olmadığını görmek için, onu Ek 1'de verilen $F_{1,2}$ dağılımının yüzdeleri ile karşılaştırırız. 90, 95 ve 99 yüzdeleri, sırasıyla, 8.5, 18.5 ve 98.5'tur. Bu nedenle 11.75'lik gözlenen değerimiz 90 yüzdeliği (%10 kritik değeri) aştığından, $\alpha = 0.10$ anlam düzeyinde sıfır hipotezini reddederiz. Yani, veri incelemeye aldığımız kitledeki dinlenmiş ve koşan insanlar arasında, ortalama olarak kalp atım hızı ile ilgili bir fark olduğu sonucunu ortaya koyar.

6.2.7. Bağımsız Örneklemeler t - Testi

B, hata vektörünün uzunluğu olmak üzere,

$$F = (\mathbf{y} \cdot \mathbf{U}_2)^2 / \left\{ \left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + (\mathbf{y} \cdot \mathbf{U}_4)^2 \right] / 2 \right\} \quad (6.20)$$

test istatistiğimizin, Bölüm 3'deki notasyonu kullanarak, $(A^2 / (B^2 / 2))$ olarak yeniden yazılabildiğine dikkat ediniz. Karekökleri alarak, eş değer

$$t = \frac{\mathbf{y} \cdot \mathbf{U}_2}{\left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + (\mathbf{y} \cdot \mathbf{U}_4)^2 \right] / 2} = \frac{A}{B / \sqrt{2}} = \frac{23 / \sqrt{4}}{\sqrt{(18+4.5) / 2}} = 3.42 \quad (6.21)$$

test istatistiğine de ulaşabiliriz.

Önceden ifade edildiği gibi, eğer $\mu_1 = \mu_2$ ise $\mathbf{y} \cdot \mathbf{U}_2$, $\mathbf{y} \cdot \mathbf{U}_3$ ve $\mathbf{y} \cdot \mathbf{U}_4$ izdüşüm uzunluklarının hepsi bir $N(0, \sigma^2)$ dağılımından gelir. Bu nedenle

$t = (\mathbf{y} \cdot \mathbf{U}_2) / \sqrt{[(\mathbf{y} \cdot \mathbf{U}_3)^2 + (\mathbf{y} \cdot \mathbf{U}_4)^2] / 2}$, Ek 1'de tanımlanan t_2 dağılımından gelir.

3.42'lik gözlenen t değerimizin büyük veya küçük olup olmadığını görmek için, onu Ek 1'de verilen t - dağılımının yüzdeleri ile karşılaştırırız. İki yanlı dağılımın 95, 97.5 ve 99.5 yüzdeleri sırasıyla bu değerlerin negatifi olan 5, 2.5 ve 0.5 yüzdeleri ile beraber 2.920, 4.303 ve 9.925'tir. O, sırasıyla -2.920 ve 2.920'lik 5 ve 95 yüzdeleriyle sınırlanan %90 normal aralıkta olmadığından 3.42'lik gözlem değerimiz büyüktür. Bu nedenle kitle ortalamalarının aynı olduğu düşüncesini reddetmemiz için yeterli bir kanıtı sahip olduğumuz sonucunu çıkarırız.

6.2.8. Yorumlar

Yukarıdaki analizde, \mathbf{U}_1 doğrultusunu iki kitle ortalamasının μ ortalamasını tahmin etmek için kullanılır, \mathbf{U}_2 doğrultusu iki kitle ortalamaları arasındaki $\mu_1 - \mu_2$ farkını tahmin etmek için kullanılır. \mathbf{U}_3 ve \mathbf{U}_4 doğrultuları σ^2 kitle varyansını tahmin etmek için kullanılır. En sonunda $\mathbf{y} \cdot \mathbf{U}_1 = \sqrt{4}\bar{y}$ izdüşüm uzunluğunun ortalaması $\sqrt{4}\mu$ olur, bu nedenle $\bar{y} = 39.75$ tahminimiz olarak işe yarar. Aynı şekilde $\mathbf{y} \cdot \mathbf{U}_2 = \bar{y}_1 - \bar{y}_2$ izdüşüm uzunluğunun ortalaması $\mu_2 - \mu_1$ olur, bu nedenle $\bar{y}_2 - \bar{y}_1 = 45.5 - 34 = 7.5$; $\mu_2 - \mu_1$ 'nin tahminimiz olarak işe yarar. Aynı zamanda $\mathbf{y} \cdot \mathbf{U}_3$ ve $\mathbf{y} \cdot \mathbf{U}_4$ izdüşüm uzunlukları karesinin her ikisinin de ortalaması σ^2 olur, bu nedenle $s^2 = [(\mathbf{y} \cdot \mathbf{U}_3)^2 + (\mathbf{y} \cdot \mathbf{U}_4)^2] / 2 = (18 + 4.5) / 2 = 11.25$; σ^2 en iyi tahminimiz olarak iş görür (Burada örneklem ortalaması için \bar{y}_1 ve \bar{y}_2 yeni notasyonunu ve genel ortalama için \bar{y} notasyonunu kullandık). Bundan başka, uzun vadede $(\mathbf{y} \cdot \mathbf{U}_2)^2$ nin ortalaması $(\mu_2 - \mu_1)^2 + \sigma^2$ iken $(\mathbf{y} \cdot \mathbf{U}_1)^2$ izdüşüm uzunluğu karesi ortalaması $4\mu^2 + \sigma^2$ 'dir. Bu nedenle eğer $\mu_1 \neq \mu_2$ ise $(\mathbf{y} \cdot \mathbf{U}_2)^2$ niceliği şişirilmiştir ve test istatistiğimiz σ^2 şişirilmiş bir tahminin σ^2 'nin bir yansız tahminiyle bölünmesi olarak düşünülebilir.

7. BASİT REGRESYON

Bu bölümde iki x ve y değişkenleri arasında en basit ilişki ile, yani $y = \alpha + \beta x$ düz doğru ilişkisiyle, ilgileniriz. Asıl ilgimiz β eğiminin sıfır olup olmadığını test etme de olacak. Bölüm 7.1’de, Christchurch, Yeni Zelanda’da kış akşamlarının bir kümesi için y hava kirliliği seviyeleri ve x inversiyon(ters) etkilerinden oluşan bir veri kümesini açıklayacağız (Saville ve Wood, 1996). Bir düz doğru uydurmak için yöntem, Bölüm 7.2’de bir veri kümesinin bir altkümesini analiz ederek uygulamaya başlanır. Bu, Bölüm 7.3’teki tam veri kümesinin bir analizi ile ilerletilir.

7.1. HAVA KİRLİLİĞİ

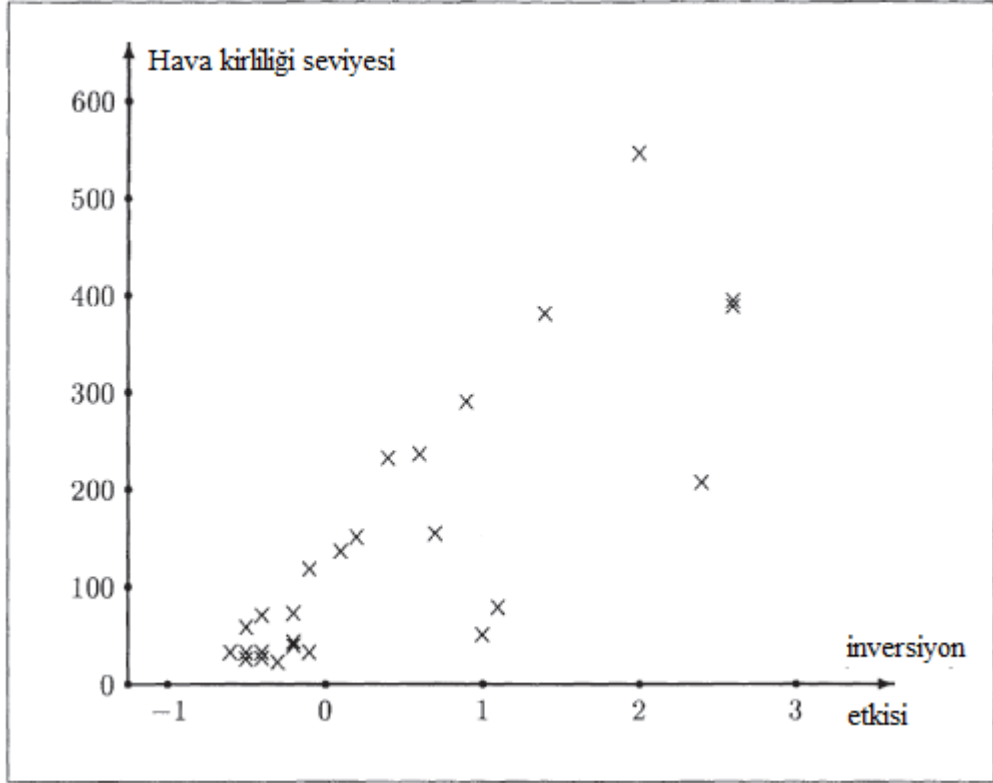
1992 kışı esnasında, Yeni Zelanda’nın Güney adasında hidroelektrik depolama göllerindeki bir su noksanlığı Christchurch’ün Güney adasında katı-yakıtların normal kullanımından daha fazlasına kullanımına sebep oldu. Ortaya çıkan hava kirliliği akşamın geç vaktinde en yüksek seviyesindeydi ve bir “inversiyon etkisi” meydana geldiğinde, (bu, hava sıcaklığının yeryüzünden yukarıda, yeryüzü seviyesinden daha yüksek olduğunda, yani, normal duruma ters olduğunda ortaya çıkar), hava kirliliğinin, akşamları çok şiddetli olacağı düşünüldü. Normal bir akşamda herhangi bir kirliliği yayan sıcak hava yükselir, bununla beraber, bir inversiyon etkisi ortaya çıktığında hava yükselmez, bu nedenle kirlilik akşam boyunca artar. Güney yarıkürede kış olan Haziran, 1992’nin 30 gününün her biri için Çizelge 7.1’de sunulan dört ana değişken, bu ilişkiyi araştırmak için bize imkan sağlayacak. Birinci değişken St.Albans, Christchurch’ün dış mahallesindeki (varoşundaki) bir izleme istasyonunda her akşam öğleden sonra 11.10’da kaydedilen hava kirliliği seviyesi veya daha kesin olarak metre küp başına mikrogramlarda askıda duran katı parçacığın seviyesidir (Saville ve Wood, 1996).

Çizelge 7.1. Hava kirliliği seviyeleri, yer seviyesinde ve yerin yukarısında sıcaklıklar ve onların farkı (inversiyon etkisi)⁶

Haziran(gün)	Hava kirliliği	Sıcaklık (°C)		İnversiyon etkisi
		Yerin yukarısında	Yerde	
1
2	290	6.0	5.1	0.9
3	79	8.9	7.8	1.1
4	33	5.7	6.3	-0.6
5	27	6.6	7.0	-0.4
6	59	6.6	7.1	-0.5
7	26	7.6	8.1	-0.5
8	33	8.5	9.0	-0.5
9	23	9.5	9.8	-0.3
10	118	4.7	4.8	-0.1
11	33	10.0	10.1	-0.1
12	394	5.6	3.0	2.6
13	73	6.2	6.4	-0.2
14	281	2.6	1.2	1.4
15	41	4.8	5.0	-0.2
16	51	8.3	7.3	0.1
17	42	5.5	5.7	-0.2
18	136	4.1	4.0	0.1
19	232	1.0	0.6	0.4
20	546	1.2	-0.8	2.0
21
22
23
24	33	4.2	4.6	-0.4
25	236	1.7	1.1	0.6
26	154	2.8	2.1	0.7
27	388	2.1	-0.5	2.6
28	151	3.4	3.2	0.2
29	71	2.7	3.1	-0.4
30	207	1.6	-0.8	2.4

⁶ Veri Mr.BobAyrey, CanterburyReginalCoincil, müsaadesiyle alındı. Eksik değerler . ile gösterilirler (Saville ve Wood, 1996).

İkinci ve üçüncü değişkenler sırasıyla, yer seviyesinde ve 3 metre yerden yukarıda hava kirliliği seviyesi ile birlikte eş anlı olarak kaydedilen hava sıcaklıklarıdır. Dördüncü değişken, iki sıcaklık arasındaki fark(yerin yukarısındaki sıcaklık eksi yer sıcaklığı) olan, inversiyon etkisinin bir ölçüsüdür. İlişkinin bir resimsel görünümü için, Şekil 7.1’te inversiyon etkisine karşı hava kirliliği seviyesinin grafiğini inceleyiniz.



Şekil 7.1. 1992, kışı haziran ayı için inversiyon etkisine karşı hava kirliliği seviyesinin saçılım diyagramı

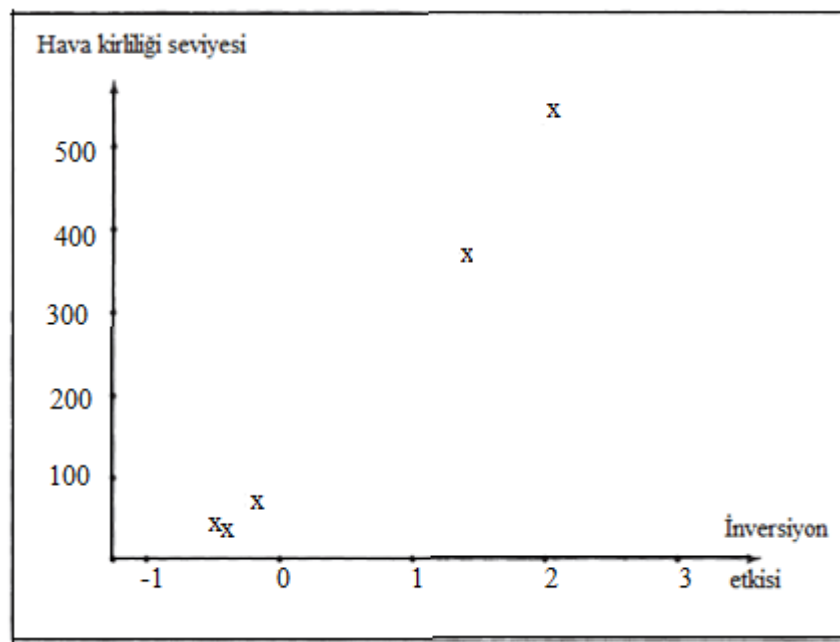
Bir “normal” akşam vakti 11.10’da yer sıcaklığı yerin yukarısındaki sıcaklıktan azıcık daha yüksektir. Bu nedenle grafiğin sol altında görüldüğü gibi, x değerleri negatiftirler.

7.2. Beş Büyüklüğüne Sahip Örneklem

Şimdi böyle veriye bir düz doğruyu uydurma problemine, geometrik yöntemimizi uygulayacağız. Açıklamanın kolaylığı için başlangıç olarak düşük boyutta bir örneklem ile çalışmayı tercih ediyoruz. Bu nedenle, Çizelge 7.1’den beş günü, yani, 13, 5, 20, 8 ve 14 Haziran,1992 günlerini seçeriz. Azaltılmış veri kümesi Çizelge 6.2’de ve azaltılmış saçılım grafiği Şekil 7.2’de gösterilmiştir.

Çizelge 7.2. Başlangıç ile ilgili örneğimiz için kullanılan 5 günlük örneklem

Haziran (gün)	Hava kirliliği	Sıcaklık (°c)		İnversiyon etkisi
		Yerin yukarısında	Yerde	
13	73	6.2	6.4	-0.2
5	27	6.6	7.0	-0.4
20	546	1.2	-0.8	2.0
8	33	8.5	9.0	-0.5
14	381	2.6	1.2	1.4



Şekil 7.2. 5 günlük örnekleminiz için inversiyon etkisine karşı hava kirliliği seviyesinin saçılım diyagramı

Sonuçta oluşan gözlem vektörü ve x değerlerinin karşılık gelen vektörü:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 73 \\ 27 \\ 546 \\ 33 \\ 381 \end{bmatrix} \quad \text{ve} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -0.2 \\ -0.4 \\ 2.0 \\ -0.5 \\ 1.4 \end{bmatrix} \quad (7.1)$$

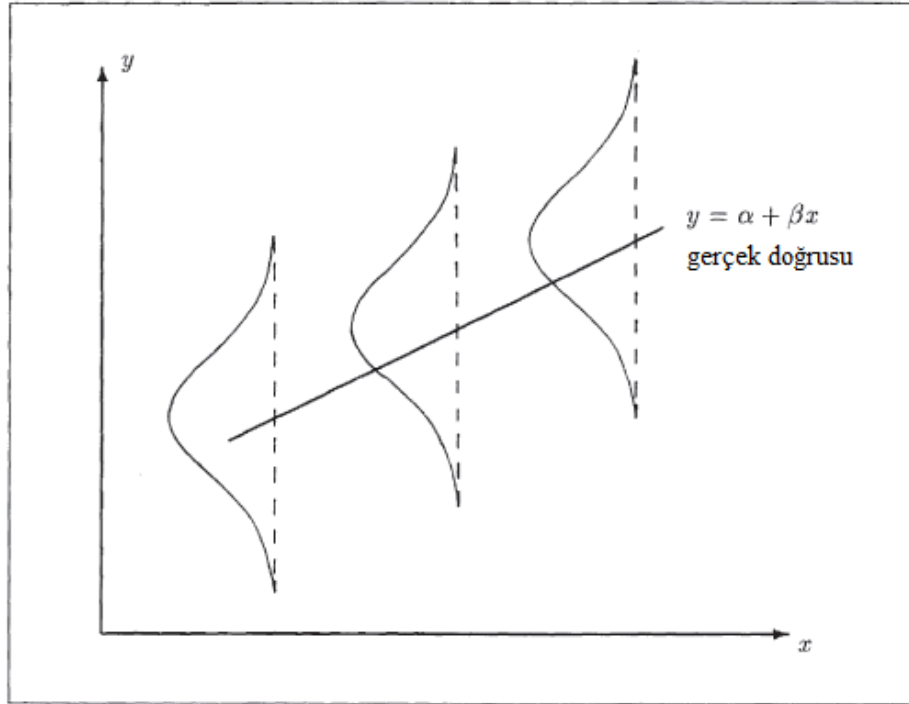
tür.

7.2.1. Amaç

İncelemenin amacı hava kirliliğinin inversiyon etkisine bağlı olup olmadığını araştırmaktır.

7.2.2. Model

İlk olarak basit(lineer-doğrusal) regresyon modeli, α sabit(regresyon sabiti) ve β eğim olmak üzere, x ve y değişkenleri arasında ilişkinin bir düz doğru, yani, $y = \alpha + \beta x$, olduğunu kabul eder. İkinci olarak, basit regresyon modeli, x değerlerinin hatasız ölçüldüklerini de kabul eder. Üçüncü olarak, model, x 'in verilen bir değeri için, y için tüm mümkün değerlerin, tüm x değerleri için aynı olan bir σ^2 varyansı ile $y = \alpha + \beta x$ doğrusu üzerindeki nokta etrafında normal dağıldıklarını kabul eder. Bu durum Şekil 7.3'te gösterilir.



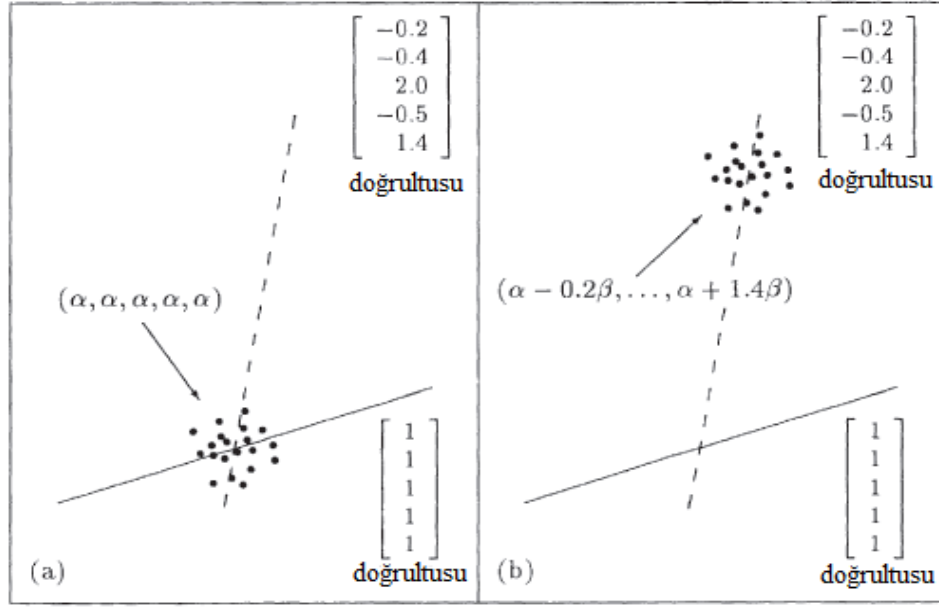
Şekil 7.3. Basit regresyon modelinin varsayımları

7.2.3. Temel Düşünce

Şekil 7.4 en son durumumuz için temel düşünceyi açıklar. Eğer hava kirliliği seviyesi ve inversiyon etkisi ilişkisiz iseler, bu takdirde, $\beta = 0$ 'dır ve gözlem vektörümüz Şekil 7.4(a)'da gösterildiği gibi, $(\alpha, \alpha, \alpha, \alpha, \alpha)$ noktası üzerinde merkezlenen bir dağılımın parçası olacaktır. Bununla beraber, eğer hava kirliliği seviyesi ile inversiyon etkisi ilişkili ise, bu takdirde $\beta \neq 0$ ve gözlem vektörümüz Şekil 7.4(b)'de gösterildiği gibi,

$$\begin{bmatrix} \alpha + \beta x_1 \\ \alpha + \beta x_2 \\ \alpha + \beta x_3 \\ \alpha + \beta x_4 \\ \alpha + \beta x_5 \end{bmatrix} = \begin{bmatrix} \alpha - 0.2\beta \\ \alpha - 0.4\beta \\ \alpha + 2.0\beta \\ \alpha - 0.5\beta \\ \alpha + 1.4\beta \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} -0.2 \\ -0.4 \\ 2.0 \\ -0.5 \\ 1.4 \end{bmatrix} \quad (7.2)$$

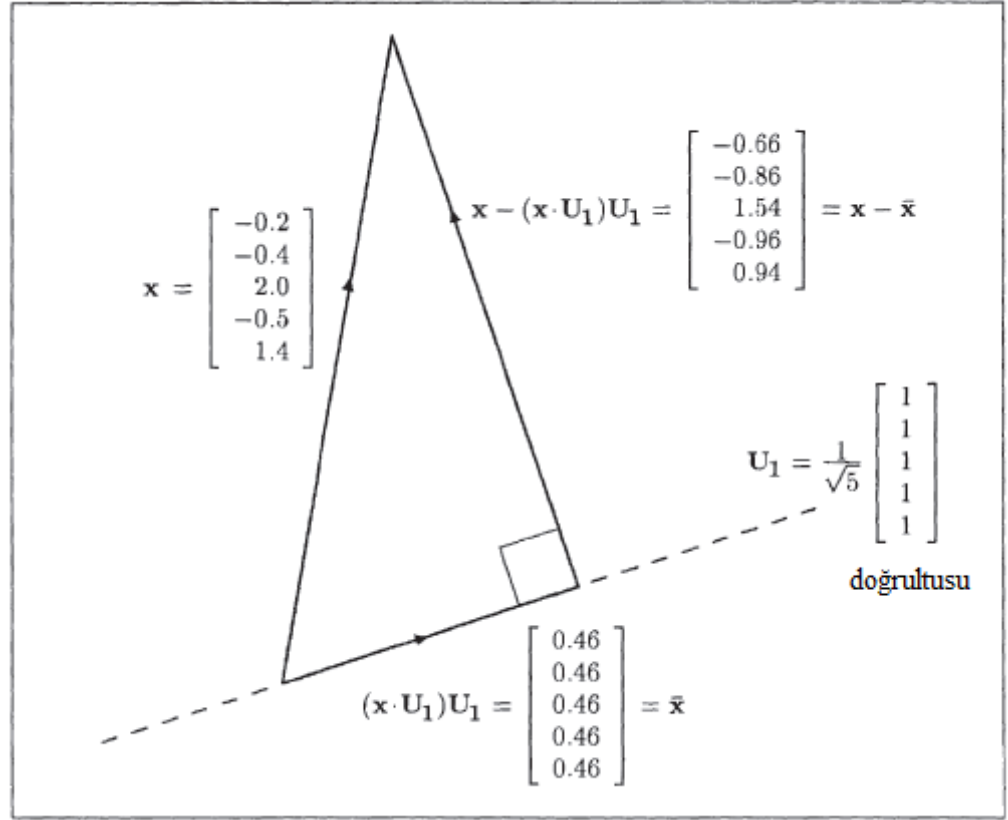
vektörünün uç noktası üzerinde merkezlenen bir dağılımın parçası olacaktır. Başka bir deyişle, eğer $\beta = 0$ ise dağılım eş açılı doğru üzerinde merkezlenir, halbuki $\beta \neq 0$ ise, dağılım $[1, 1, 1, 1, 1]'$ ve $[-0.2, -0.4, 2.0, -0.5, 1.4]'$ vektörleri tarafından tanımlanan düzlemde eş açılı doğrudan uzaklaşır. İncelememizin birçok tekrarlama düşünülduğünde aynı beş x değeri ile incelemeleri düşünüyor olduğumuza dikkat ediniz. Yani, Şekil 7.4'te, örneklem noktalarındaki değişme, sadece x için beş sabit değere karşılık gelen y değerlerindeki olası değişmeden dolayıdır.



Şekil 7.4. Hava kirliliği incelememizin birçok tekrarlanmasını temsil eden veri noktalarının dağılımları

7.2.4. Uygun Bir Koordinat Sistemi

Modeli standart tarzımızda uydurabilmek için, $[1, 1, 1, 1, 1]'$ ve $[-0.2, -0.4, 2.0, -0.5, 1.4]'$ vektörleri tarafından tanımlanan düzlem içinde uzanan iki ortogonal doğrultuyu içeren 5-boyutlu uzay için bir ortogonal koordinat eksenleri takımına ihtiyaç duyarız. İlk olarak $U_1 = [1, 1, 1, 1, 1]'/\sqrt{5}$ alabiliriz. Bu yukarıdaki düzlemde uzanan ve U_1 'e dik olan, ikinci bir U_2 birim vektörüne ihtiyacımız olduğunu söyler. Böyle bir vektörü nasıl buluruz? Şekil 7.5 ihtiyacımız olan ipuçlarını verir. Şekilde ilk olarak, $U_1 = [1, 1, 1, 1, 1]'/\sqrt{5}$ doğrultusunu ve $x = [-0.2, -0.4, 2.0, -0.5, 1.4]'$ vektörünü çizeriz. Bu ikisi ilgilendiğimiz düzlemi belirler.



Şekil 7.5. U_1 ve x değerlerinin vektörü tarafından tanımlanan düzlemde $(x - \bar{x})$ doğrultusu

Bundan sonra, x vektörünü U_1 doğrultusu üzerine izdüşürerek

$$(\mathbf{x} \cdot \mathbf{U}_1) \mathbf{U}_1 = \begin{pmatrix} -0.2 \\ -0.4 \\ 2.0 \\ -0.5 \\ 1.4 \end{pmatrix} \cdot \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \cdot \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = 0.46 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \bar{\mathbf{x}} \quad (7.3)$$

izdüşüm vektörünü elde ederiz. Daha sonra, Şekil 7.5'te gösterildiği gibi, x değerlerinin vektöründen bu $\bar{\mathbf{x}}$ izdüşüm vektörünü çıkarırız. Böylece,

$$\mathbf{x} - (\mathbf{x} \cdot \mathbf{U}_1) \mathbf{U}_1 = \begin{pmatrix} -0.2 \\ -0.4 \\ 2.0 \\ -0.5 \\ 1.4 \end{pmatrix} - \begin{pmatrix} 0.46 \\ 0.46 \\ 0.46 \\ 0.46 \\ 0.46 \end{pmatrix} = \begin{pmatrix} -0.66 \\ -0.86 \\ 1.54 \\ -0.96 \\ 0.94 \end{pmatrix} = \mathbf{x} - \bar{\mathbf{x}} \quad (7.4)$$

vektörünü elde ederiz. Sonuçta oluşan $\mathbf{x} - \bar{\mathbf{x}}$ vektörü, \mathbf{U}_1 ortogonal ve $[1, 1, 1, 1, 1]'$ ve $[-0.2, -0.4, 2.0, -0.5, 1.4]'$ vektörleri tarafından belirlenen düzlemde uzanır. Böylece, geriye sadece onu, onun uzunluğuna, yani, $\|\mathbf{x} - \bar{\mathbf{x}}\|$ 'ye, bölerek birim vektör yapmak kalır. $5.32 = (-0.66)^2 + (-0.86)^2 + 1.54^2 + (-0.96)^2 + 0.94^2$, yani, $(x_i - \bar{x})$ değerlerinin karelerinin toplamı olmak üzere, bu vektörünü verir.

$$\mathbf{U}_2 = \frac{\mathbf{x} - \bar{\mathbf{x}}}{\|\mathbf{x} - \bar{\mathbf{x}}\|} = \frac{1}{\sqrt{5.352}} \begin{bmatrix} -0.66 \\ -0.86 \\ 1.54 \\ -0.96 \\ 0.94 \end{bmatrix} \quad (7.5)$$

5-boyutlu uzay için sonuçta oluşan ortogonal koordinat sistemi

$$\mathbf{U}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{U}_2 = \frac{1}{\sqrt{5.352}} \begin{bmatrix} -0.66 \\ -0.86 \\ 1.54 \\ -0.96 \\ 0.94 \end{bmatrix}, \mathbf{U}_3, \mathbf{U}_4 \text{ ve } \mathbf{U}_5 \quad (7.6)$$

dir. Burada $\mathbf{U}_3, \mathbf{U}_4$ ve \mathbf{U}_5 "hata uzayı" doğrultularıdır. Bunları açık bir şekilde yazmak önceki bölümlerdeki kadar kolay değildir, bu nedenle bu, ilerlememiz için gerekli olmadığından, öyle yapmayacağız.

7.2.5. Ortogonal Ayrışım

$\mathbf{y} = [73, 27, 546, 33, 381]'$ gözlem vektörümüzü ilk iki koordinat eksenleri, yani \mathbf{U}_1 ve \mathbf{U}_2 üzerine izdüşürürüz. Birinci izdüşüm vektörü

$$(\mathbf{y} \cdot \mathbf{U}_1) \mathbf{U}_1 = \left(\begin{bmatrix} 73 \\ 27 \\ 546 \\ 33 \\ 381 \end{bmatrix} \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right) \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 212 \\ 212 \\ 212 \\ 212 \\ 212 \end{bmatrix} = \bar{\mathbf{y}} \quad (7.7)$$

ve ikinci izdüşüm vektörü,

$$\begin{aligned}
 (\mathbf{y} \cdot \mathbf{U}_2) \mathbf{U}_2 &= \begin{pmatrix} 73 \\ 27 \\ 546 \\ 33 \\ 381 \end{pmatrix} \cdot \frac{1}{\sqrt{5.352}} \begin{pmatrix} -0.66 \\ -0.86 \\ 1.54 \\ -0.96 \\ 0.94 \end{pmatrix} \cdot \frac{1}{\sqrt{5.352}} \begin{pmatrix} -0.66 \\ -0.86 \\ 1.54 \\ -0.96 \\ 0.94 \end{pmatrix} \\
 &= \frac{73x(-0.66) + 27x(-0.86) + 546x1.54 + 33x(-0.96) + 381x0.94}{\sqrt{5.352}x\sqrt{5.352}} \begin{pmatrix} -0.66 \\ -0.86 \\ 1.54 \\ -0.96 \\ 0.94 \end{pmatrix} \\
 &= \frac{1095.6}{5.352} \begin{pmatrix} -0.66 \\ -0.86 \\ 1.54 \\ -0.96 \\ 0.94 \end{pmatrix} = 204.765 \begin{pmatrix} -0.66 \\ -0.86 \\ 1.54 \\ -0.96 \\ 0.94 \end{pmatrix} = b(\mathbf{x} - \bar{\mathbf{x}}) = \begin{pmatrix} -135.145 \\ -176.098 \\ 315.338 \\ -196.574 \\ 192.479 \end{pmatrix} \quad (7.8)
 \end{aligned}$$

dir. Burada 204.765 tahmin edilen b eğimidir. Yukarıdaki hesaplamalara yeniden geri dönüp bakarak, tahmin edilen eğim için cebirsel formülün

$$b = \frac{y_1x(x_1 - \bar{x}) + \dots + y_5x(x_5 - \bar{x})}{(x_1 - \bar{x})^2 + \dots + (x_5 - \bar{x})^2} = \frac{1095.9}{5.352} = 204.765 \quad (7.9)$$

olduğunu görebiliriz. Gözlem vektörünün sonuçta oluşan ortogonal ayrışımı:

$$\mathbf{y} = (\mathbf{y} \cdot \mathbf{U}_1) \mathbf{U}_1 + (\mathbf{y} \cdot \mathbf{U}_2) \mathbf{U}_2 + (\mathbf{y} \cdot \mathbf{U}_3) \mathbf{U}_3 + (\mathbf{y} \cdot \mathbf{U}_4) \mathbf{U}_4 + (\mathbf{y} \cdot \mathbf{U}_5) \mathbf{U}_5 \quad (7.10)$$

$$\begin{pmatrix} 73 \\ 27 \\ 546 \\ 33 \\ 381 \end{pmatrix} = \begin{pmatrix} 212 \\ 212 \\ 212 \\ 212 \\ 212 \end{pmatrix} + \begin{pmatrix} -135.1 \\ -176.1 \\ 315.3 \\ -196.6 \\ 192.5 \end{pmatrix} + \begin{pmatrix} -3.9 \\ -8.9 \\ 18.7 \\ 17.6 \\ -23.5 \end{pmatrix} \quad (7.11)$$

$$\begin{array}{cccc}
 \mathbf{y} & = & \bar{\mathbf{y}} & + & b(\mathbf{x} - \bar{\mathbf{x}}) & + & [\mathbf{y} - \bar{\mathbf{y}} - b(\mathbf{x} - \bar{\mathbf{x}})] \\
 \text{gözlem} & & \text{ortalama} & & \text{eğim} & & \text{hata} \\
 \text{vektörü} & & \text{vektörü} & & \text{vektörü} & & \text{vektörü}
 \end{array} \quad (7.12)$$

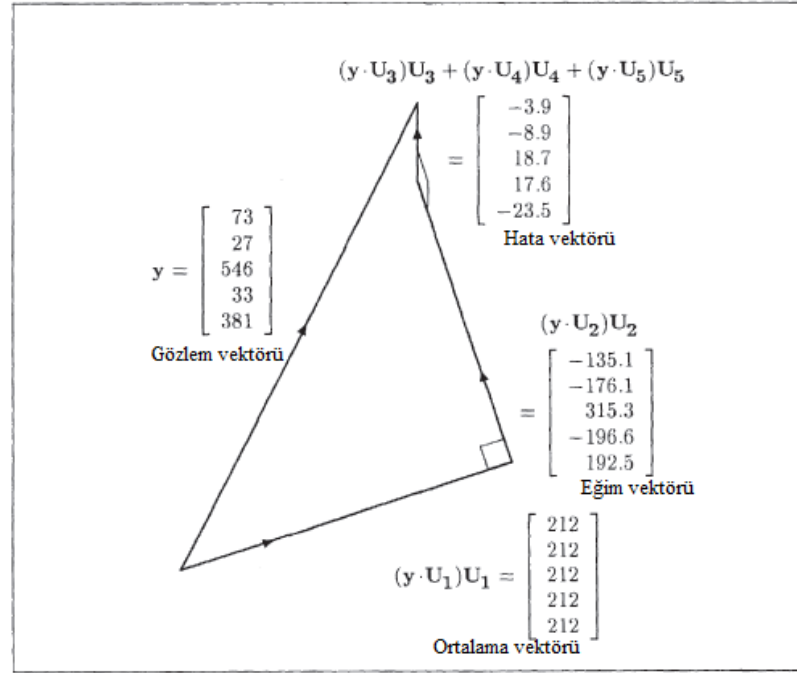
dir. Burada hata vektörü çıkarmayla elde edilir. Özet olarak, y gözlem vektörünü, Şekil 7.6'da gösterildiği gibi, bir ortalama vektörü, bir eğim vektörü ve bir hata vektörü olmak üzere üç ortogonal bileşene ayırdık. Uydurduğumuz doğrunun denklemi

$$y = 212 + 204.8(x - 0.46) \quad (7.13)$$

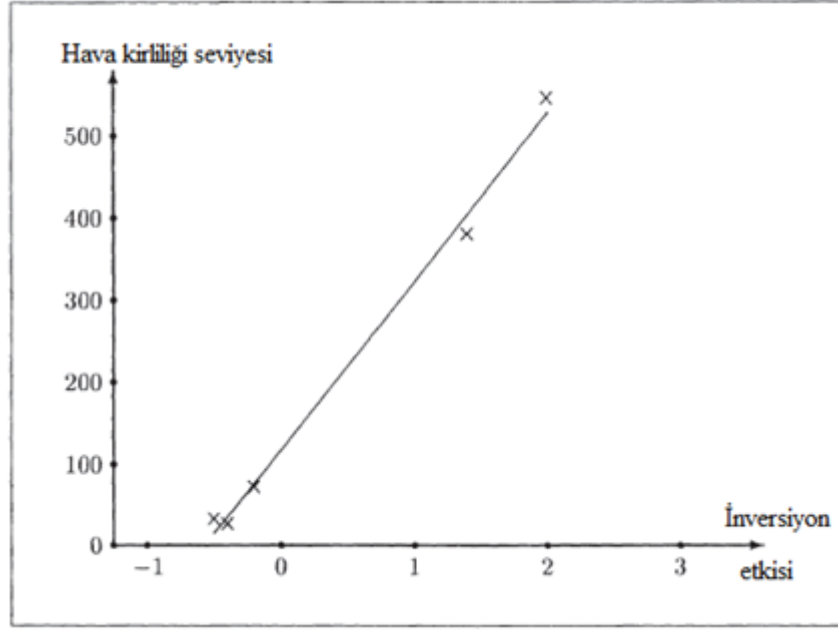
dır. Sonuçta oluşan uygun doğru, Şekil 7.7'de saçılım diyagramı ile birlikte, gösterilir. Genel olarak, denklemin,

$$y = \bar{y} + b(x - \bar{x}) \quad (7.14)$$

ortogonal biçimde olduğuna dikkat ediniz.



Şekil 7.6. Basit regresyon durumu için gözlem vektörünün ortogonal ayrışımı



Şekil 7.7. İncersiyon etkisine karşı hava kirliliđinin sađılım diyagramı üzerinde birleřtirilmiř uygun regresyon

Parantezi ađarak ve terimleri toplayarak denklemini orijinal, ortogonal olmayan, biđime geri evrilebiliriz. Bylece,

$$y = 212 + 204.8x - 204.8 \times 0.46$$

$$= (212 - 204.8 \times 0.46) + 204.8x \quad (7.15)$$

$$y = 117.8 + 204.8x$$

elde ederiz. Genel olarak, ortogonal olmayan biđim

$$y = (\bar{y} - b\bar{x}) + bx \quad (7.16)$$

dir. Bu nedenle α sabit terimi, $\alpha = \bar{y} - b\bar{x}$ (burada o 117.8'dir.) ve n rneklem byklđ olmak zere, β eđimini

$$b = \frac{y_1x_1(x_1 - \bar{x}) + \dots + y_nx_n(x_n - \bar{x})}{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2} \quad (7.17)$$

ile tahmin edilir.

7.2.6. İzdüşüm Uzunluklarının Dağılımları

Gerçek β eğiminin makul bir biçimde sıfır olup olmadığını test edilmesine hazırlanmak için şimdi $\mathbf{y} \cdot \mathbf{U}_1, \mathbf{y} \cdot \mathbf{U}_2, \mathbf{y} \cdot \mathbf{U}_3, \mathbf{y} \cdot \mathbf{U}_4$ ve $\mathbf{y} \cdot \mathbf{U}_5$ izdüşüm uzunluklarının dağılımlarının ortalamalarını hesaplarız. Bu uygulamayı daha sonra tekrar yapmaktan tasarruf etmek için y_1, y_2 ve x_1, x_2 özel değerlerini kullanmanın yerine örneğimizde x gibi cebirsel sembolleri kullanarak, oldukça genel terimler ile ilgileniriz.

İlk olarak,

$$\mathbf{y} \cdot \mathbf{U}_1 = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{\sqrt{5}}, \text{dir.} \quad (7.18)$$

Bu,

$$\begin{aligned} &= \frac{(\alpha + \beta x_1) + (\alpha + \beta x_2) + (\alpha + \beta x_3) + (\alpha + \beta x_4) + (\alpha + \beta x_5)}{\sqrt{5}} \\ &= \frac{5\alpha + \beta(x_1 + x_2 + x_3 + x_4 + x_5)}{\sqrt{5}} = \frac{5\alpha + 5\beta \bar{x}}{\sqrt{5}} = \sqrt{5}(\alpha + \beta \bar{x}) \end{aligned} \quad (7.19)$$

lik bir ortalama değere sahiptir. İkinci olarak,

$$\mathbf{y} \cdot \mathbf{U}_2 = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \cdot \frac{1}{\|\mathbf{x} - \bar{\mathbf{x}}\|} \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \\ x_4 - \bar{x} \\ x_5 - \bar{x} \end{bmatrix} = \frac{y_1(x_1 - \bar{x}) + \dots + y_5(x_5 - \bar{x})}{\|\mathbf{x} - \bar{\mathbf{x}}\|} \quad (7.20)$$

dir. Bu,

$$\begin{aligned}
&= \frac{(\alpha + \beta x_1)(x_1 - \bar{x}) + (\alpha + \beta x_2)(x_2 - \bar{x}) + \dots + (\alpha + \beta x_5)(x_5 - \bar{x})}{\|\mathbf{x} - \bar{\mathbf{x}}\|} \\
&= \frac{\alpha(x_1 + x_2 + x_3 + x_4 + x_5 - 5\bar{x}) + \beta[(x_1(x_1 - \bar{x}) + \dots + x_5(x_5 - \bar{x}))]}{\|\mathbf{x} - \bar{\mathbf{x}}\|} \\
&= \frac{0 + \beta[(x_1(x_1 - \bar{x}) + \dots + x_5(x_5 - \bar{x}))] - \beta\bar{x}[(x_1 - \bar{x}) + \dots + (x_5 - \bar{x})]}{\|\mathbf{x} - \bar{\mathbf{x}}\|} \quad (7.21)
\end{aligned}$$

lik bir ortalama değere sahiptir. Burada paya bir sıfır üçüncü vektörü ekledik.

$$\begin{aligned}
&= \frac{\beta[(x_1 - \bar{x})^2 + \dots + (x_5 - \bar{x})^2]}{\sqrt{(x_1 - \bar{x})^2 + \dots + (x_5 - \bar{x})^2}} \\
&= \beta \sqrt{(x_1 - \bar{x})^2 + \dots + (x_5 - \bar{x})^2} = \beta \|\mathbf{x} - \bar{\mathbf{x}}\| \quad (7.22)
\end{aligned}$$

dır. Son olarak, hata uzayındaki açıkça belirtilmemiş koordinat doğrultularının, yani, $\mathbf{y} \cdot \mathbf{U}_3$, $\mathbf{y} \cdot \mathbf{U}_4$ ve $\mathbf{y} \cdot \mathbf{U}_5$ 'in, her birisi için izdüşüm uzunluğunun ortalamasının sıfır olduğunu göstermemiz gerekir. Bunu $\mathbf{U}_e = [e_1, e_2, e_3, e_4, e_5]'$ gibi, genel bir doğrultu için yaparız. Şimdi,

$$\mathbf{y} \cdot \mathbf{U}_e = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \cdot \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix} = y_1 e_1 + y_2 e_2 + y_3 e_3 + y_4 e_4 + y_5 e_5 \quad (7.23)$$

dir. Bu,

$$\begin{aligned}
&= [\alpha + \beta x_1]e_1 + [\alpha + \beta x_2]e_2 + [\alpha + \beta x_3]e_3 + [\alpha + \beta x_4]e_4 + [\alpha + \beta x_5]e_5 \\
&= [(\alpha + \beta\bar{x}) + \beta(x_1 - \bar{x})]e_1 + \dots + [(\alpha + \beta\bar{x}) + \beta(x_5 - \bar{x})]e_5 \quad (7.24)
\end{aligned}$$

dir. Burada her bir terimde $\beta\bar{x}$ yi ekledik ve bir de çıkardık.

$$= (\alpha + \beta \bar{x})(e_1 + \dots + e_5) + \beta[(x_1 - \bar{x})e_1 + \dots + (x_5 - \bar{x})e_5]$$

$$= (\alpha + \beta \bar{x}) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix} + \beta \begin{bmatrix} x_1 - \bar{x} \\ x_1 - \bar{x} \\ x_1 - \bar{x} \\ x_1 - \bar{x} \\ x_1 - \bar{x} \end{bmatrix} \cdot \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix}$$

$$= 0 \quad (7.25)$$

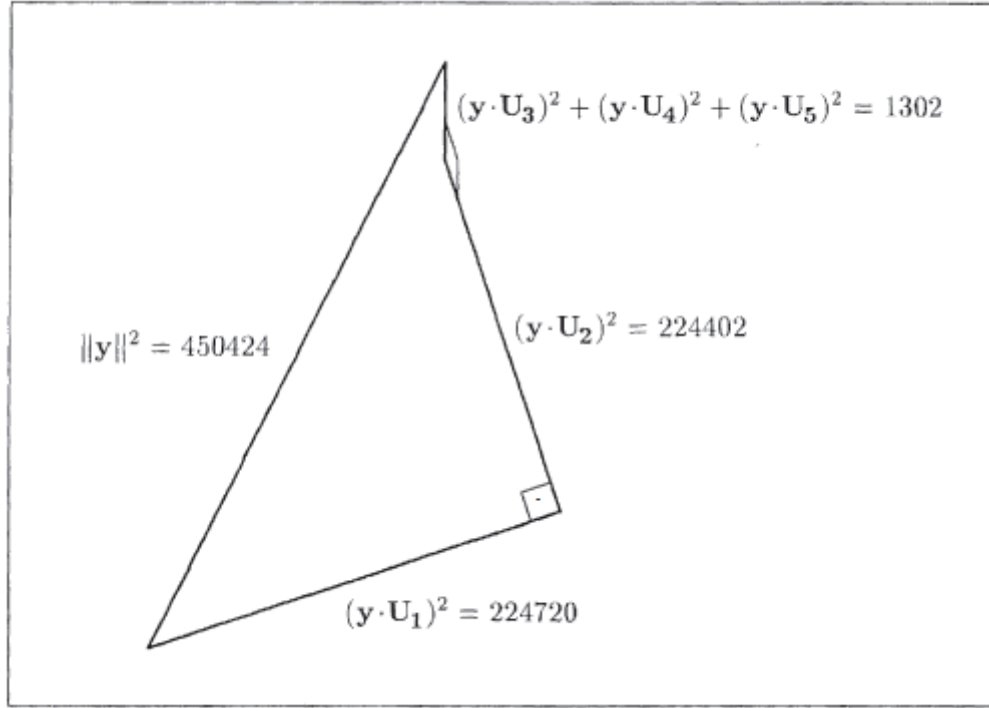
dır. Çünkü, \mathbf{U}_e , hem \mathbf{U}_1 ve hem de \mathbf{U}_2 'ye ortogonaldır. Ayrıca, beş izdüşüm uzunluğunun tümünün σ^2 varyansına sahip olduğu kolaylıkla gösterilebilir. Bu nedenle $\mathbf{y} \cdot \mathbf{U}_1$ bir $N(\sqrt{5}(\alpha + \beta \bar{x}), \sigma^2)$ dağılımından, $\mathbf{y} \cdot \mathbf{U}_2$ bir $N(\beta \|\mathbf{x} - \bar{\mathbf{x}}\|, \sigma^2)$ dağılımından ve $\mathbf{y} \cdot \mathbf{U}_3, \mathbf{y} \cdot \mathbf{U}_4$ ve $\mathbf{y} \cdot \mathbf{U}_5$ bir $N(0, \sigma^2)$ dağılımından gelir. Dikkat edilmesi gereken nokta $\mathbf{y} \cdot \mathbf{U}_3, \mathbf{y} \cdot \mathbf{U}_4$ ve $\mathbf{y} \cdot \mathbf{U}_5$ 'in dağılımları daima sıfır üzerinde merkezlenirken, $\mathbf{y} \cdot \mathbf{U}_1$ ve $\mathbf{y} \cdot \mathbf{U}_2$ 'nin dağılımları potansiyel olarak, sıfırdan farklı nicelikler üzerinde merkezlendikleridir. Bundan başka, β 'nin sıfır veya sıfırdan farklı olmadığına karar vermek için \mathbf{U}_2 doğrultusu ilgili doğrultudur, çünkü eğer $\beta = 0$ ise $\mathbf{y} \cdot \mathbf{U}_2$ izdüşüm uzunluğu sıfır ortalamasını alır, eğer $\beta \neq 0$ ise, $\mathbf{y} \cdot \mathbf{U}_2$ izdüşüm uzunluğu ortalama olarak sıfırdan farklı bir niceliği alır.

7.2.7. Hipotezi Test Etme

Hava kirliliği seviyesi inversiyon etkisine bağlı mıdır? Basit regresyon modelimize göre bu, gerçek β eğimi = 0 mıdır? sorusuna çevrilir. Konuyla ilgili Pisagor parçalanması Şekil 7.8'de gösterildiği gibi,

$$\|\mathbf{y}\|^2 = (\mathbf{y} \cdot \mathbf{U}_1)^2 + (\mathbf{y} \cdot \mathbf{U}_2)^2 + (\mathbf{y} \cdot \mathbf{U}_3)^2 + (\mathbf{y} \cdot \mathbf{U}_4)^2 + (\mathbf{y} \cdot \mathbf{U}_5)^2$$

$$450424 = 224720 + 224402 + 1302 \quad (7.26)$$



Şekil 7.8. Basit regresyon örneğimiz için, gözlem vektörünün uzunluğunun karesinin, beş ortogonal izdüşümün kareleri toplamı olarak Pisagor parçalanması

$\beta = 0$ olup olmadığını test etmek için $(\mathbf{y} \cdot \mathbf{U}_2)^2$ izdüşüm uzunluğu karesinin $(\mathbf{y} \cdot \mathbf{U}_3)^2$, $(\mathbf{y} \cdot \mathbf{U}_4)^2$ ve $(\mathbf{y} \cdot \mathbf{U}_5)^2$ izdüşüm uzunlukları karelerinin ortalamasının aynısı, veya ondan önemli derece daha büyük olup olmadığını kontrol ederiz. Sonuçta oluşan test istatistiği

$$F = \frac{(\mathbf{y} \cdot \mathbf{U}_2)^2}{\left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + (\mathbf{y} \cdot \mathbf{U}_4)^2 + (\mathbf{y} \cdot \mathbf{U}_5)^2 \right] / 3} = \frac{224402}{1302 / 3} = \frac{224402}{434} = 517 \quad (7.27)$$

dir. Bu gözlenen F değeri büyük veya küçük müdür? Şimdi, eğer $\beta = 0$ ise, $\mathbf{y} \cdot \mathbf{U}_2, \mathbf{y} \cdot \mathbf{U}_3, \mathbf{y} \cdot \mathbf{U}_4$ ve $\mathbf{y} \cdot \mathbf{U}_5$ uzunluklarının tümü bir $N(0, \sigma^2)$ dağılımından gelirler, bu nedenle,

$$F = (\mathbf{y} \cdot \mathbf{U}_2)^2 / \left\{ \left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + (\mathbf{y} \cdot \mathbf{U}_4)^2 + (\mathbf{y} \cdot \mathbf{U}_5)^2 \right] / 3 \right\} \quad (7.28)$$

Ek 1'de tanımlanan, $F_{1,3}$ dağılımından gelir. Bundan dolayı, 517'lik gözlenen değerimizin büyük veya küçük olup olmadığını görmek için, onu, Ek 1'de verilen

$F_{1,3}$ dağılımının yüzdeleri ile karşılaştırırız. 90,95 ve 99 yüzdeleri, sırasıyla 5.54, 10.13 ve 34.12'dir, bu nedenle 517'lik gözlenen değerimiz açık olarak 34.12'lik 99 yüzdeliğinden çok daha büyüktür. Yani, hava kirliliğinin inversiyon etkisine bağlı olduğu düşüncesine desteklemek için çok güçlü bir delil elde ettik(yani, $\beta \neq 0$).

7.2.8. Eşdeğer t -Testi

$F = (\mathbf{y} \cdot \mathbf{U}_2)^2 / \left\{ \left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + (\mathbf{y} \cdot \mathbf{U}_4)^2 + (\mathbf{y} \cdot \mathbf{U}_5)^2 \right] / 3 \right\}$ test istatistiğimiziz, B ; hata vektörünün uzunluğu olmak üzere $A^2 / (B^2 / 3)$ olarak yeniden yazılabilir. Karekök alarak, bu, Bölüm 3'te verilen

$$t = \frac{(\mathbf{y} \cdot \mathbf{U}_2)^2}{\sqrt{\left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + (\mathbf{y} \cdot \mathbf{U}_4)^2 + (\mathbf{y} \cdot \mathbf{U}_5)^2 \right] / 3}} = \frac{A}{B / \sqrt{3}} = \frac{\sqrt{224402}}{\sqrt{1302/3}} = 22.74 \quad (7.29)$$

biçimini alır. Bundan sonra; yukarıda elde edilene denk bir sonucu veren, bu t değeri Ek 1'de verilen t_3 dağılımının yüzdeleri ile karşılaştırılır.

7.2.9. Yorumlar

Yukarıdaki analizde, \mathbf{U}_1 doğrultusu ($\alpha + \beta \bar{x}$) sabitini tahmin etmek için kullanılır, \mathbf{U}_2 doğrultusu gerçek β eğimini tahmin etmek için kullanılır ve $\mathbf{U}_3, \mathbf{U}_4$ ve \mathbf{U}_5 doğrultuları σ^2 varyansını tahmin etmek için kullanılır. Uzun vade de, $\mathbf{y} \cdot \mathbf{U}_2 = b \|\mathbf{x} - \bar{\mathbf{x}}\|$ izdüşüm uzunluğu $\beta \|\mathbf{x} - \bar{\mathbf{x}}\|$ ortalamasını alır, bundan dolayı, $b = 204.8$ β 'nin bir tahmini olarak hizmet görür. İzdüşüm uzunlukları karelerinin tümü $(\mathbf{y} \cdot \mathbf{U}_3)^2, (\mathbf{y} \cdot \mathbf{U}_4)^2$ ve $(\mathbf{y} \cdot \mathbf{U}_5)^2$ σ^2 ortalamasını alırlar, bu nedenle, $s^2 = \left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + (\mathbf{y} \cdot \mathbf{U}_4)^2 + (\mathbf{y} \cdot \mathbf{U}_5)^2 \right] / 3 = 1302 / 3 = 434$ 'te σ^2 'nin tahmini olarak hizmet görür. Ayrıca, uzun vade de, $(\mathbf{y} \cdot \mathbf{U}_2)^2$ izdüşüm uzunluğu karesi $\beta^2 \|\mathbf{x} - \bar{\mathbf{x}}\|^2 + \sigma^2$ ortalamasını alır. Bu nedenler, eğer $\beta \neq 0$ ise, niceliği büyüktür(şişkindir) ve $(\mathbf{y} \cdot \mathbf{U}_2)^2 / s^2$ test istatistiğimiziz, σ^2 'nin bir şişkin tahmini olarak düşünülebilir.

7.3. Genel Durum

Genel durumu açıklamak için, şimdi geometrik yaklaşımımızdaki temel adımları belirler ve Çizelge 7.1’de verilen “26 atlanmış olmayan” günün tam kümesine onları uygularız. Sonuçta oluşan gözlem vektörü ve x değerlerinin karşılık gelen vektörü

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_{26} \end{bmatrix} = \begin{bmatrix} 290 \\ \vdots \\ 207 \end{bmatrix} \text{ ve } \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{26} \end{bmatrix} = \begin{bmatrix} 0.9 \\ \vdots \\ 2.4 \end{bmatrix} \quad (7.30)$$

tür.

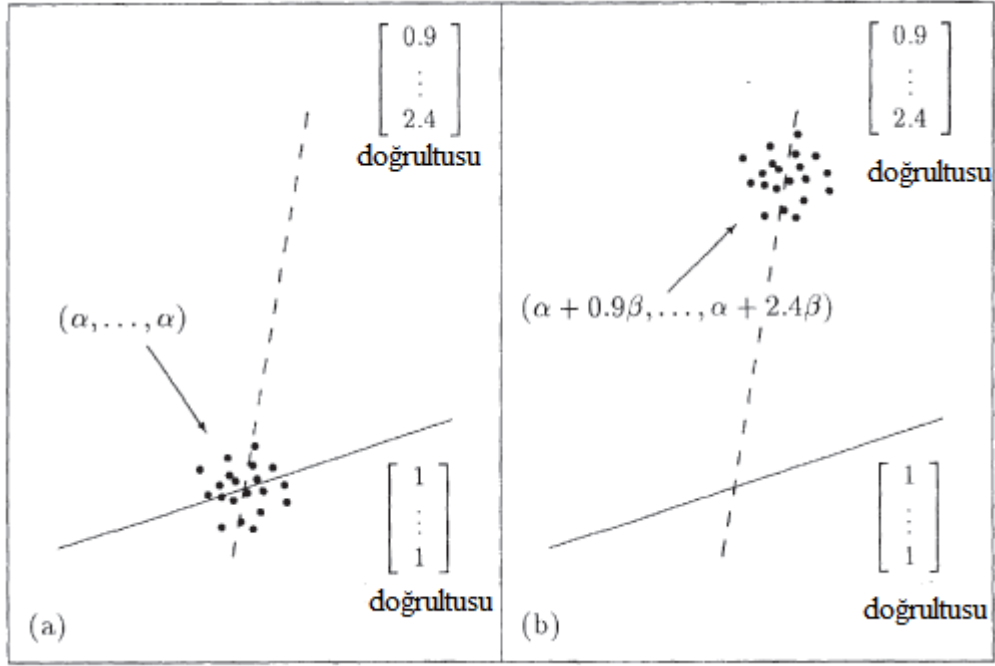
7.3.1. Temel Düşünce

Temel düşünce Bölüm 7.2’deki aynıdır. Eğer hava kirliliği seviyesi ve inversiyon etkisi ilişkisiz iseler, bu takdirde $\beta = 0$ ’dır ve gözlem vektörümüz Şekil 7.9(a)’da gösterildiği gibi, (α, \dots, α) noktası üzerinde merkezlenen bir dağılımın (saçılımın) parçası olacak. Bununla beraber eğer, hava kirliliği seviyesi ve inversiyon etkisi ilişkili iseler, bu takdirde $\beta \neq 0$ olur ve gözlem vektörümüz, Şekil 7.9(b)’de gösterildiği gibi

$$\begin{bmatrix} \alpha + \beta x_1 \\ \vdots \\ \alpha + \beta x_{26} \end{bmatrix} = \begin{bmatrix} \alpha + 0.9\beta \\ \vdots \\ \alpha + 2.4\beta \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 0.9 \\ \vdots \\ 2.4 \end{bmatrix} \quad (7.31)$$

vektörünün uç noktası üzerinde merkezlenen bir dağılımın parçası olacak. Bir başka deyişle, eğer $\beta = 0$ ise, dağılım eşaçılı doğru üzerinde merkezlenir, halbuki, $\beta \neq 0$ ise dağılım $[1, \dots, 1]'$ ve $[0.9, \dots, 2.4]'$ vektörleri tarafından tanımlanan (belirlenen) düzlemde eş açılı doğrultudan uzaklaşır.

Bölüm 7.2’deki gibi, incelememizin birçok tekrarlama düşünülürken, incelememizi aynı 26 tane x değeriyle düşünüyorduk şimdiye kadar. Yani, Şekil 7.9’da örneklem noktalarındaki değişim büyük olasılıkla sadece x için 26 tane sabit değere karşılık gelen y değerlerindeki değişim nedeniyledir.



Şekil 7.9. Kirlilik incelememizin birçok tekrarlamasının temsil eden veri noktalarının dağılımları

7.3.2. Uygun Bir Koordinat Düzlemi

26- boyutlu uzay için uygun bir koordinat sistemi

$$\mathbf{U}_1 = \frac{1}{\sqrt{26}} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \mathbf{U}_2 = \frac{1}{\sqrt{25.9646}} \begin{bmatrix} 0.454 \\ \vdots \\ 1.946 \end{bmatrix}, \mathbf{U}_3, \dots, \mathbf{U}_{26} \quad (7.32)$$

dır. Burada \mathbf{U}_1 alışlagelen eş açılı doğrultudur ve \mathbf{U}_2 , x değerlerinin onların $\bar{x} = 0.446$ ortalaması etrafında $(x_i - \bar{x})$ sapmalarının vektörünün onun $\|\mathbf{x} - \bar{\mathbf{x}}\| = \sqrt{25.9646}$ uzunluğuna bölünmesinden oluşan vektördür. $\mathbf{U}_3, \dots, \mathbf{U}_{26}$ doğrultuları 24 ortogonal “hata uzayı” doğrultusudur.

7.3.3. Ortogonal Ayrışım

$\mathbf{y} = [290, \dots, 207]'$ gözlem vektörümüzü ilk iki koordinat eksenine üzerine, yani, \mathbf{U}_1 ve \mathbf{U}_2 koordinat eksenleri üzerine, izdüşürürüz. Birinci izdüşüm vektörü,

$$(\mathbf{y} \cdot \mathbf{U}_1) \mathbf{U}_1 = \left(\begin{bmatrix} 290 \\ \vdots \\ 207 \end{bmatrix} \cdot \frac{1}{\sqrt{26}} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right) \frac{1}{\sqrt{26}} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 148.3 \\ \vdots \\ 148.3 \end{bmatrix} = \bar{\mathbf{y}} \quad (7.33)$$

ve ikinci izdüşüm vektörü

$$\begin{aligned} &= (\mathbf{y} \cdot \mathbf{U}_2) \mathbf{U}_2 = \left(\begin{bmatrix} 290 \\ \cdot \\ \cdot \\ \cdot \\ 207 \end{bmatrix} \cdot \frac{1}{\sqrt{25.9646}} \begin{bmatrix} 0.454 \\ \cdot \\ \cdot \\ \cdot \\ 1.954 \end{bmatrix} \right) \frac{1}{\sqrt{25.9646}} \begin{bmatrix} 0.454 \\ \cdot \\ \cdot \\ \cdot \\ 1.954 \end{bmatrix} \\ &= \frac{290 \times 0.454 + \dots + 207 \times 1.954}{\sqrt{25.9646} \cdot \sqrt{25.9646}} \begin{bmatrix} 0.454 \\ \cdot \\ \cdot \\ \cdot \\ 1.954 \end{bmatrix} \\ &= \frac{3035.1}{25.9646} \begin{bmatrix} 0.454 \\ \cdot \\ \cdot \\ \cdot \\ 1.954 \end{bmatrix} = 116.9 \begin{bmatrix} 0.454 \\ \cdot \\ \cdot \\ \cdot \\ 1.954 \end{bmatrix} = b(\mathbf{x} - \bar{\mathbf{x}}) = \begin{bmatrix} 53.1 \\ \cdot \\ \cdot \\ \cdot \\ 228.4 \end{bmatrix} \quad (7.34) \end{aligned}$$

dir. Burada tahmin edilen eğim = 116.9'dur. Gözlem vektörünün sonuçta oluşan ortogonal ayrışımı, hata vektörünü çıkarmayla elde edilmek üzere,

$$\mathbf{y} = (\mathbf{y} \cdot \mathbf{U}_1) \mathbf{U}_1 + (\mathbf{y} \cdot \mathbf{U}_2) \mathbf{U}_2 + (\mathbf{y} \cdot \mathbf{U}_3) \mathbf{U}_3 + \dots + (\mathbf{y} \cdot \mathbf{U}_{26}) \mathbf{U}_{26}$$

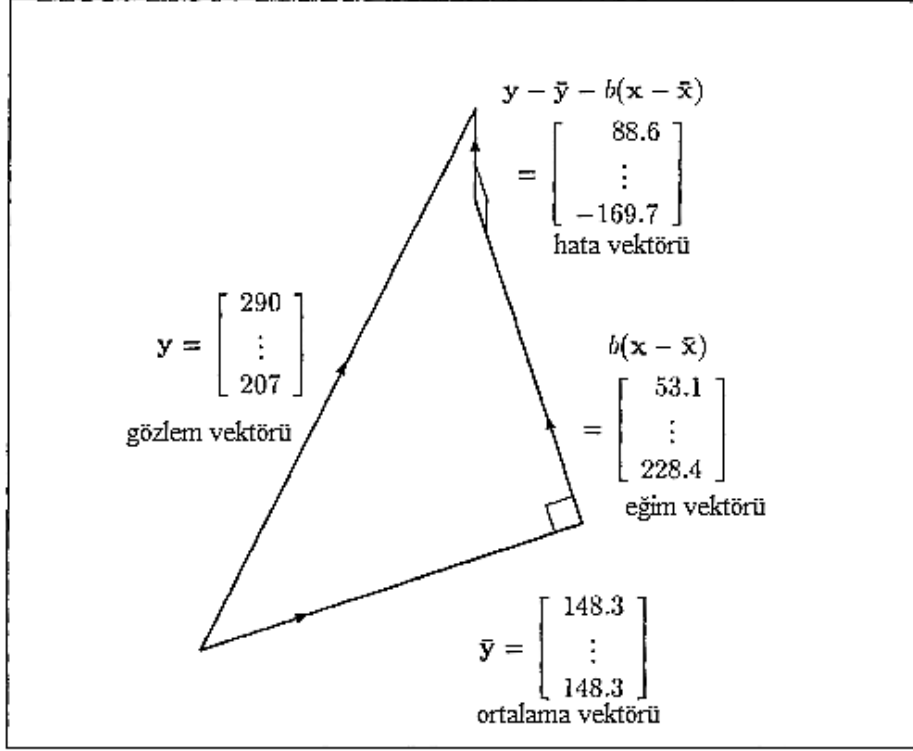
$$\begin{bmatrix} 209 \\ \vdots \\ 207 \end{bmatrix} = \begin{bmatrix} 148.3 \\ \vdots \\ 148.3 \end{bmatrix} + \begin{bmatrix} 53.1 \\ \vdots \\ 228.4 \end{bmatrix} + \begin{bmatrix} 88.6 \\ \vdots \\ -169.7 \end{bmatrix} \quad (7.35)$$

$$\mathbf{y} = \bar{\mathbf{y}} + b(\mathbf{x} - \bar{\mathbf{x}}) + [\mathbf{y} - \bar{\mathbf{y}} - b(\mathbf{x} - \bar{\mathbf{x}})]$$

gözlem ortalama eğim hata
vektörü vektörü vektörü vektörü

$$(7.36)$$

dır. Bölüm 7.2'deki gibi, \mathbf{y} gözlem vektörümüzü, Şekil 7.10'da gösterildiği gibi, biri ortalama vektörü, biri eğim vektörü ve biri hata vektörü olmak üzere üç ortogonal bileşene ayırdık.



Şekil 7.10. Çizelge 6.1'den tam veri kümesini kullanan, basit regresyon için gözlem vektörünün ortogonal ayrışımı

Ortogonal formda uydurduğumuz doğrultman denklemi

$$y = 148.3 + 116.9(x - 0.446) \quad (7.37)$$

dır. Sonuçta oluşan uygun doğru, saçılım diyagramı ile birlikte, Şekil 7.11'de gösterilir. Ortogonal olmayan forma çevrildiğinde denklem

$$y = 148.3 + 116.9x - 116.9x \cdot 0.446$$

$$y = (148.3 - 116.9 \cdot 0.446) + 116.9x \quad (7.38)$$

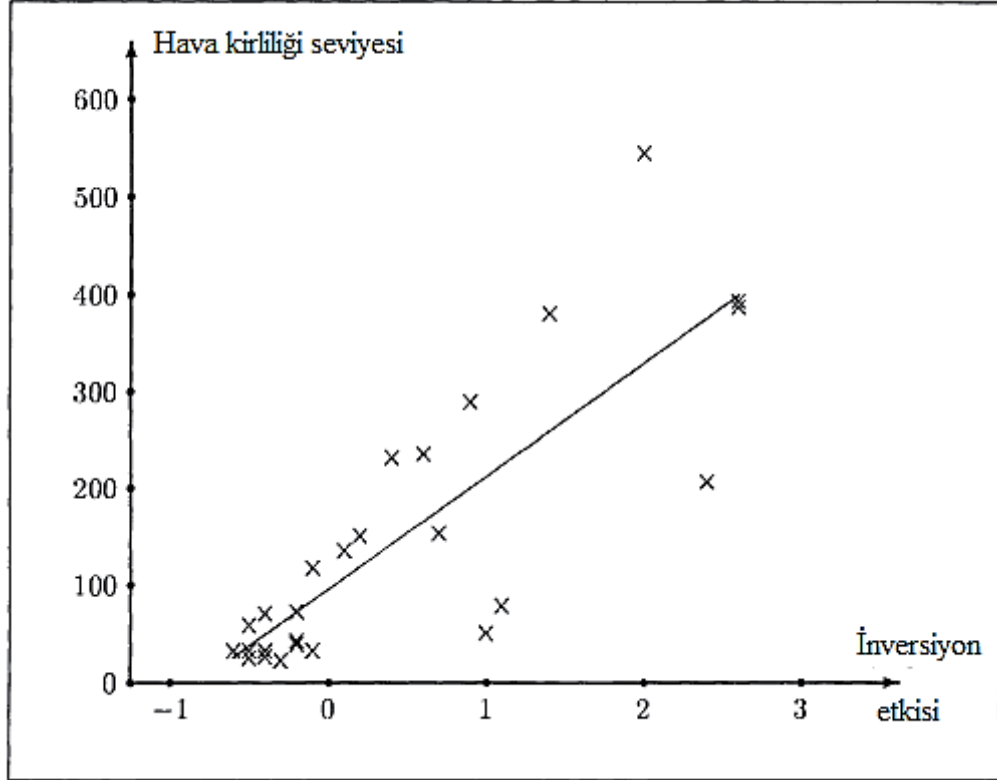
veya

$$y = 96.2 + 116.9x \quad (7.39)$$

olur. Bölüm 7.2’de belirtildiği gibi, α sabit terimi $a = y - b\bar{x} = 96.2$ ve n (burada 26) örneklem büyüklüğü olmak üzere, β eğimi

$$b = \frac{\mathbf{y}(\mathbf{x}_1 - \bar{\mathbf{x}})}{\|\mathbf{x}_1 - \bar{\mathbf{x}}\|^2} = \frac{y_1x_1(x_1 - \bar{x}) + \dots + y_nx_n(x_n - \bar{x})}{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2} = 116.9 \quad (7.40)$$

ile tahmin edilir.



Şekil 7.11. 1992 kışı haziran ayı için inversiyon etkisine karşı hava kirliliğinin saçılım diyagramı üzerine yerleştirilen uygun regresyon doğrusu

7.3.4. İzdüşüm Uzunluklarının Dağılımları

Bölüm 7.2’deki gibi devam ederek, $\mathbf{y} \cdot \mathbf{U}_1$ in $N(\sqrt{26}(\alpha + \beta\bar{x}), \sigma^2)$ dağılımından, $\mathbf{y} \cdot \mathbf{U}_2$ ’nin bi $N(\beta\|\mathbf{x} - \bar{\mathbf{x}}\|, \sigma^2)$ dağılımından ve $\mathbf{y} \cdot \mathbf{U}_3, \dots, \mathbf{y} \cdot \mathbf{U}_{26}$ ’nın tümünün bir $N(0, \sigma^2)$ dağılımından geldiğini buluruz. Önceden olduğu gibi, belirtilmesi gereken husus $\mathbf{y} \cdot \mathbf{U}_3, \dots, \mathbf{y} \cdot \mathbf{U}_{26}$ dağılımları daima sıfır üzerinde merkezlenirken, $\mathbf{y} \cdot \mathbf{U}_1$ ve $\mathbf{y} \cdot \mathbf{U}_2$ ’nin dağılımlarının potansiyel olarak sıfırdan farklı nicelikler üzerinde merkezlenmeleridir. Bundan başka, β ’nin sıfır olup olmadığına karar vermek için

ilgili doğrultu U_2 doğrultusudur, çünkü $\beta = 0$ ise, $\mathbf{y} \cdot U_2$ 'nin ortalaması sıfırdır, ve $\beta \neq 0$ ise, $\mathbf{y} \cdot U_2$ 'nin ortalaması sıfırdan farklı bir niceliktir.

7.3.5. Hipotezi Test Etme

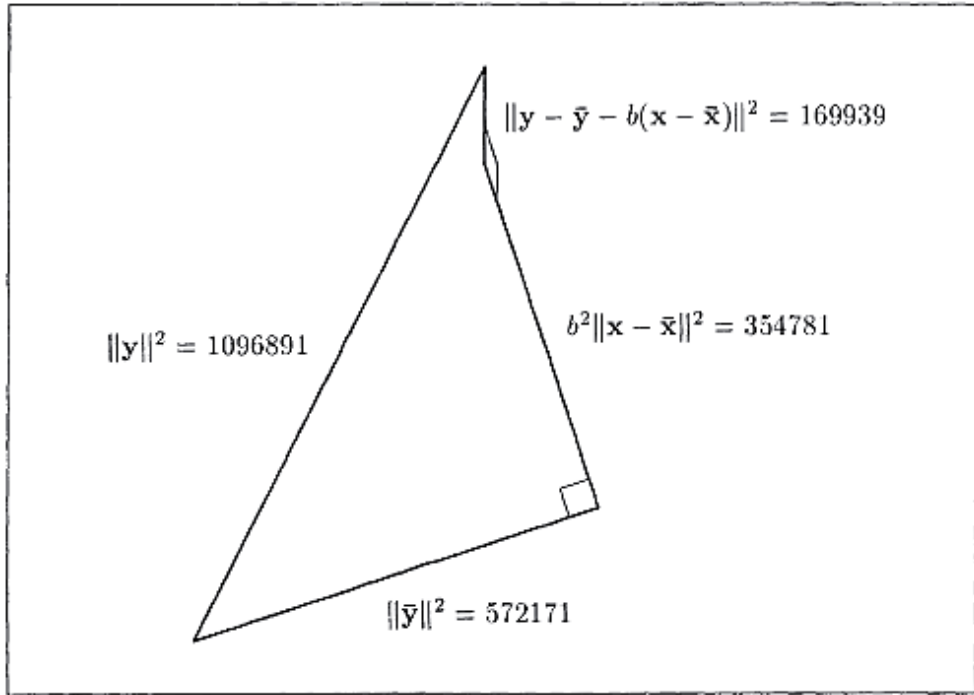
Şimdi inceleme amacımızı araştırmaya devam ederiz: Hava kirliliği seviyesi invasyon etkisine bağlı mıdır? Yani, gerçek β eğimi = 0 mıdır? Konuyla ilgili Pisagor parçalanması; Şekil 7.12'de gösterildiği gibi,

$$\|\mathbf{y}\|^2 = (\mathbf{y} \cdot U_1)^2 + (\mathbf{y} \cdot U_2)^2 + (\mathbf{y} \cdot U_3)^2 + \dots + (\mathbf{y} \cdot U_{26})^2 \quad (7.41)$$

$$\|\mathbf{y}\|^2 = \|\bar{\mathbf{y}}\|^2 + b^2 \|\mathbf{x} - \bar{\mathbf{x}}\|^2 + \|\mathbf{y} - \bar{\mathbf{y}} - b(\mathbf{x} - \bar{\mathbf{x}})\|^2$$

$$1096891 = 572171 + 354781 + 169939 \quad (7.42)$$

dir.



Şekil 7.12. Basit regresyon için gözlem vektörünün uzunluğunun karesinin Pisagor parçalanması

Pisagor parçalanmamızda $(\mathbf{y} \cdot U_2)^2$ izdüşüm uzunluğu karesine çoğu kez “regresyon kareler toplamı” olarak başvurulur. Onun cebirsel açılımı:

$$(\mathbf{y} \cdot \mathbf{U}_2)^2 = \left[\frac{\mathbf{y}(\mathbf{x} - \bar{\mathbf{x}})}{\|\mathbf{x} - \bar{\mathbf{x}}\|} \right]^2 = \frac{\left[\sum_{i=1}^n y_i (x_i - \bar{x}) \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\left[\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (7.43)$$

dır. Burada payı yeniden düzenlemek için $\sum_{i=1}^n \bar{y}(x_i - \bar{x}) = 0$ eşitliğini kullandık. x ve y 'ye göre simetrik olduğundan, ekseriyetle payın son şekli tercih edilir. $\beta = 0$ olduğu hipotezini test etmek için, $(\mathbf{y} \cdot \mathbf{U}_2)^2$ izdüşüm uzunluğu karesinin, “hata” doğrultuları için izdüşüm uzunlukları karelerinin ortalamasının, yani, $\left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + \dots + (\mathbf{y} \cdot \mathbf{U}_{26})^2 \right] / 24$ aynısı, veya ondan daha büyük olup olmadığını kontrol ederiz. Sonuçta oluşan test istatistiği

$$\begin{aligned} F &= \frac{(\mathbf{y} \cdot \mathbf{U}_2)^2}{\left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + \dots + (\mathbf{y} \cdot \mathbf{U}_{26})^2 \right] / 24} = \frac{b^2(\mathbf{x} - \bar{\mathbf{x}})^2}{\|\mathbf{y} - \bar{\mathbf{y}} - b(\mathbf{x} - \bar{\mathbf{x}})\|^2 / 24} \\ &= \frac{354781}{169939 / 24} = \frac{354781}{7081} = 50.10 \end{aligned} \quad (7.44)$$

dır. Eğer, $\beta = 0$ ise, bu; $F_{1,24}$ dağılımından gelir. Gözlenen F değeri (50.10), $F_{1,24}$ dağılımının 90 yüzdesinden daha büyük olduğundan, $\beta = 0$ olduğu hipotezini reddederiz ve hava kirliliği ve inversiyon etkisinin ilişkili olduklarını ileri sürmek için çok güçlü bir delilin var olduğu sonucunu çıkarırız.

7.3.6. Denk t -Testi

$$\begin{aligned} F &= (\mathbf{y} \cdot \mathbf{U}_2)^2 / \left\{ \left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + \dots + (\mathbf{y} \cdot \mathbf{U}_{26})^2 \right] / 24 \right\} \\ &= b^2(\mathbf{x} - \bar{\mathbf{x}})^2 / \left[\|\mathbf{y} - \bar{\mathbf{y}} - b(\mathbf{x} - \bar{\mathbf{x}})\|^2 / 24 \right] = 50.10 \end{aligned} \quad (7.45)$$

test istatistiğimizin,

$$t = \frac{(\mathbf{y} \cdot \mathbf{U}_2)^2}{\sqrt{\left[(\mathbf{y} \cdot \mathbf{U}_3)^2 + \dots + (\mathbf{y} \cdot \mathbf{U}_{26})^2 \right] / 24}}$$

$$= \frac{b(\mathbf{x} - \bar{\mathbf{x}})}{\|\mathbf{y} - \bar{\mathbf{y}} - b(\mathbf{x} - \bar{\mathbf{x}})\| / \sqrt{24}} = \frac{\sqrt{354781}}{\sqrt{169939/24}} 7.078 \quad (7.46)$$

istatistiğine dönüştürülebildiğine dikkat ediniz. Eğer $\beta = 0$ ise, bu; t_{24} dağılımından gelir. Önceki gibi, gözlenen t değerimizin büyüklüğü hava kirliliği ve inversiyon etkisinin ilişkili olmadıkları düşüncesine karşı çok güçlü delil oluşturur.

t için cebirsel ifadenin

$$t = \frac{b(\mathbf{x} - \bar{\mathbf{x}})}{s} = \frac{b\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{s} = \frac{b}{s / \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (7.47)$$

olduğuna dikkat ediniz. Burada son ifadedeki payda “eğimin standart hatası” olarak bilinir.

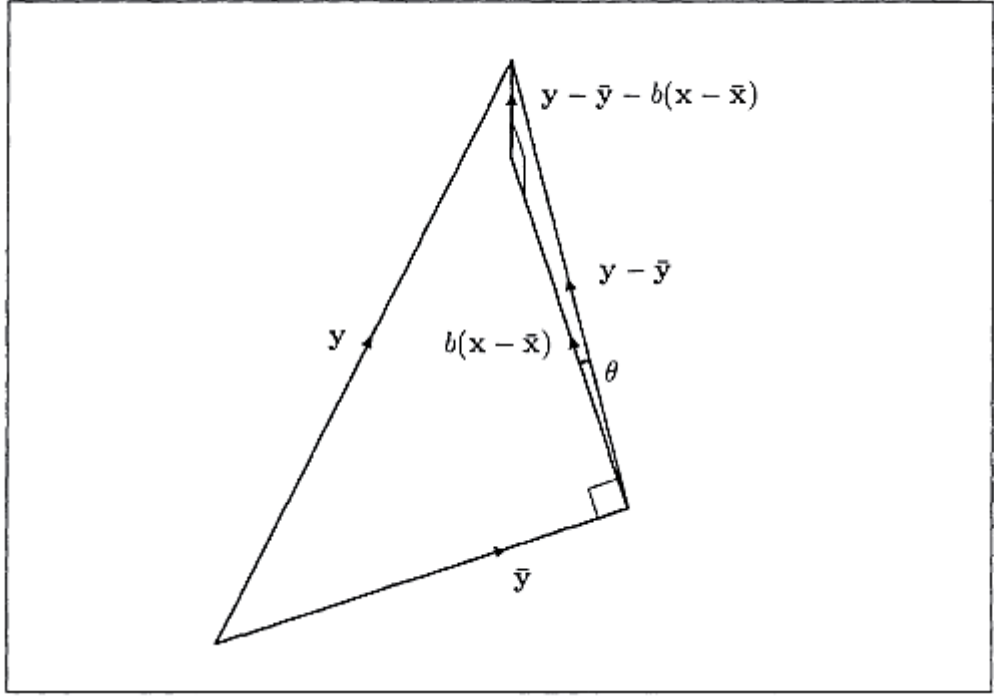
7.3.7. Korelasyon Katsayısı

β eğiminin sıfır olup olmadığını test etmenin bir üçüncü yolu vardır. Bu; r örneklem korelasyon katsayısını hesaplamak ve gözlenen değeri, $\beta = 0$ hipotezi altında r 'nin dağılımının yüzdeleri ile (Ek 1’de verilen) karşılaştırmaktır.

Korelasyon katsayısı, Şekil 7.13’te gösterildiği gibi, $(\mathbf{y} - \bar{\mathbf{y}})$ ve $(\mathbf{x} - \bar{\mathbf{x}})$ vektörleri arasında θ açısının kosinüsü olarak tanımlanabilir. (b 'nin uygun olarak pozitif olduğuna dikkat ediniz.) Örneğimizde o;

$$\begin{aligned} r = \cos \theta &= \frac{\mathbf{x} - \bar{\mathbf{x}}}{\|\mathbf{x} - \bar{\mathbf{x}}\|} \cdot \frac{\mathbf{y} - \bar{\mathbf{y}}}{\|\mathbf{y} - \bar{\mathbf{y}}\|} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \\ &= \frac{3035.1}{\sqrt{25.9646} \sqrt{524720}} = 0.822 \end{aligned} \quad (7.48)$$

olarak işe yarar. Bu değer 24 “serbestlik dereceli” korelasyon katsayısının sıfır dağılımı için Ek 1’deki 0.496 olan 99.5yüzdeden daha küçüktür, bu nedenle yine, $\beta = 0$ hipotezine karşı çok güçlü kanıtın var olduğu sonucunu çıkarırız.



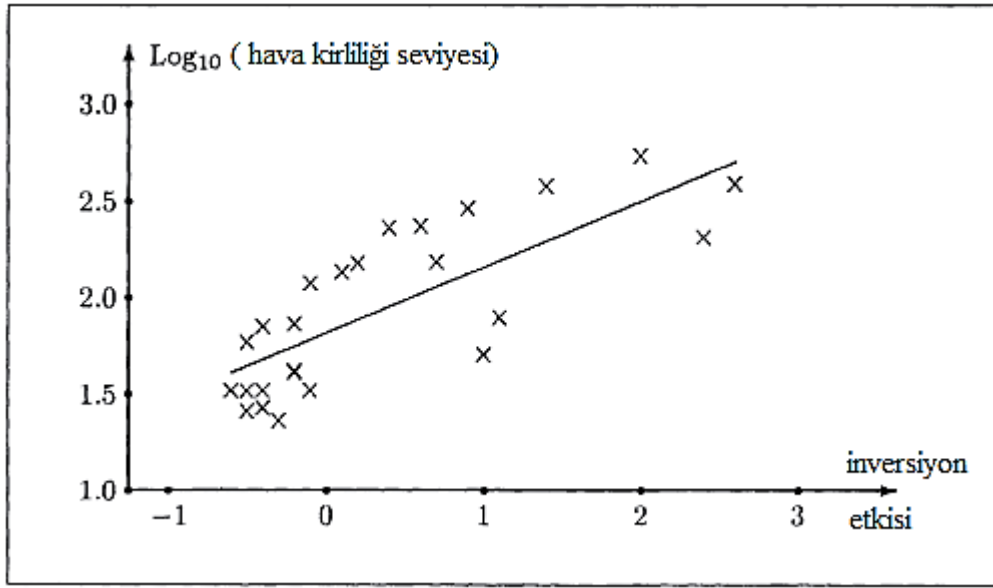
Şekil 7.13. $(y - \bar{y})$ “düzeltilmiş” gözlem vektörü ve kosinüsü r korelasyon katsayısı olan θ açısını da gösteren ortogonal ayrışımı

$t = A / (B / \sqrt{n-2}) = \sqrt{n-2} \cot \theta$ olduğundan, t değerinin θ açısının bir trigonometrik ifadesi olarak ifade edilebileceğine de dikkat ediniz.

7.3.8. Yorumlar

Bölüm 7.2’deki gibi, U_3, \dots, U_{26} “hata uzayı” doğrultuları σ^2 varyansını tahmin etmek için kullanılır iken, U_1 ve U_2 “model uzayı” doğrultuları sabit terimi ve eğimi (veya eşdeğer olarak, bir sabiti ve eğimi) tahmin etmek için kullanılır. Örneklem büyüklüğümüz 5’den 26’ya artarken, F değerimizin gerçekten azaldığını, yani, 517 civarından 50 civarına düştüğünün farkına varabiliriz. Bu uzun vadede beklediğimizin tersidir ve sadece uydurulan doğruya alışılmadık şekilde yakın uzanan beş noktamız ile, beş büyüklüğüne sahip bir örnekleme seçmedeki şansımızdan kaynaklandı. İncersiyon etkisinin yüksek hava kirliliği seviyelerine sebep olduğunu ispatlamadığımızı dikkat ediniz. Aynı ölçüde etkili olabilen üçüncü bir ilgili değişken var olabilirdi. Bizim durumumuzda, sıcaklığın daha soğuk sıcaklık ateşlerini yakan insanların sayısını daha büyüttüğünden ve hava kirliliğini daha fazla artırdığından, yer sıcaklığı böyle bir değişkendir. Aynı zamanda daha soğuk sıcaklık,

bir inversiyon etkisinin şansını daha büyütür ve hava kirliliğini daha büyütür. Yani, veri kümemizdeki soğukluk ve inversiyon etkisi yakinen ilişkilidir. Son olarak gözlenen y değerlerimizin varyansın x değerinin artışıyla arttığından Şekil 7.3'teki üçüncü varsayımın ihlal edilmiş olduğu dikkate alınmalıdır. Bizim durumumuzda, uygun gözükten bir çare y değerlerinin logaritmasını almaktır. Sonuçta oluşan noktasal diyagram ve uygun doğru Şekil 7.14'te gösterilir. Bu çizim orijinal ölçeğe geri dönüştürüldüğünde, uygun doğrunun üstel eğri olduğuna dikkat ediniz.



Şekil 7.14. Logaritmaya (10 tabanlı) dönüştürülen hava kirliliği

7.3.9. Özet

Özet olarak, basit regresyon durumunda özel öneme sahip doğrultu $U_2 = (\mathbf{x} - \bar{\mathbf{x}}) / \|\mathbf{x} - \bar{\mathbf{x}}\|$ doğrultusudur. Eğer β eğimi sıfır değilse, bu takdirde bu, izdüşüm uzunluğunun arttığı doğrultudur. Daha tam olarak, eğer $\beta \neq 0$ ise, gözlem vektörünün U_2 doğrultusu üzerine izdüşümünün uzunluğu U_3, \dots, U_n "hata" doğrultuları için izdüşüm uzunluklarına göre daha büyük olacak. Bu nedenle $\beta = 0$ ve $\beta \neq 0$ durumları arasında ayırım yapmak için U_2 doğrultusu için izdüşüm uzunluğunun karesini "hata uzayındaki" $(n-1)$ tane dik doğrultular için izdüşüm uzunluklarının ortalamasına böleriz; eğer sonuçta oluşan oran küçük ise, β 'nin sıfır olabileceğine karar veririz, halbuki bu oran büyükse β 'nin sıfır olamayacağına karar

veririz. Genel olarak, gözlemlerin vektörünün, yani $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]'$ nün $\mathbf{U}_2 = (\mathbf{x} - \bar{\mathbf{x}}) / \|\mathbf{x} - \bar{\mathbf{x}}\|$ doğrultusu üzerine izdüşümü, $b = \mathbf{y}(\mathbf{x} - \bar{\mathbf{x}}) / \|\mathbf{x} - \bar{\mathbf{x}}\|^2$ olmak üzere $b(\mathbf{x} - \bar{\mathbf{x}})$ vektörüdür. Bu yaklaşım bizi Şekil 7.15'te gösterilen $\mathbf{y} = \bar{\mathbf{y}} + b(\mathbf{x} - \bar{\mathbf{x}}) + [\mathbf{y} - \bar{\mathbf{y}} - b(\mathbf{x} - \bar{\mathbf{x}})]$ ortogonal ayrışımına götürür. Burada "hata vektörü", yani $[\mathbf{y} - \bar{\mathbf{y}} - b(\mathbf{x} - \bar{\mathbf{x}})]$, hata uzayını geren ortogonal koordinat eksenlerinin her biri için bir tane olan, $(n-2)$ tane ortogonal izdüşüm vektörünün bir toplamıdır. Sonuçta oluşan Pisagor parçalanması

$$\|\mathbf{y}\|^2 = \|\bar{\mathbf{y}}\|^2 + b^2 \|\mathbf{x} - \bar{\mathbf{x}}\|^2 + \|\mathbf{y} - \bar{\mathbf{y}} - b(\mathbf{x} - \bar{\mathbf{x}})\|^2 \quad (7.49)$$

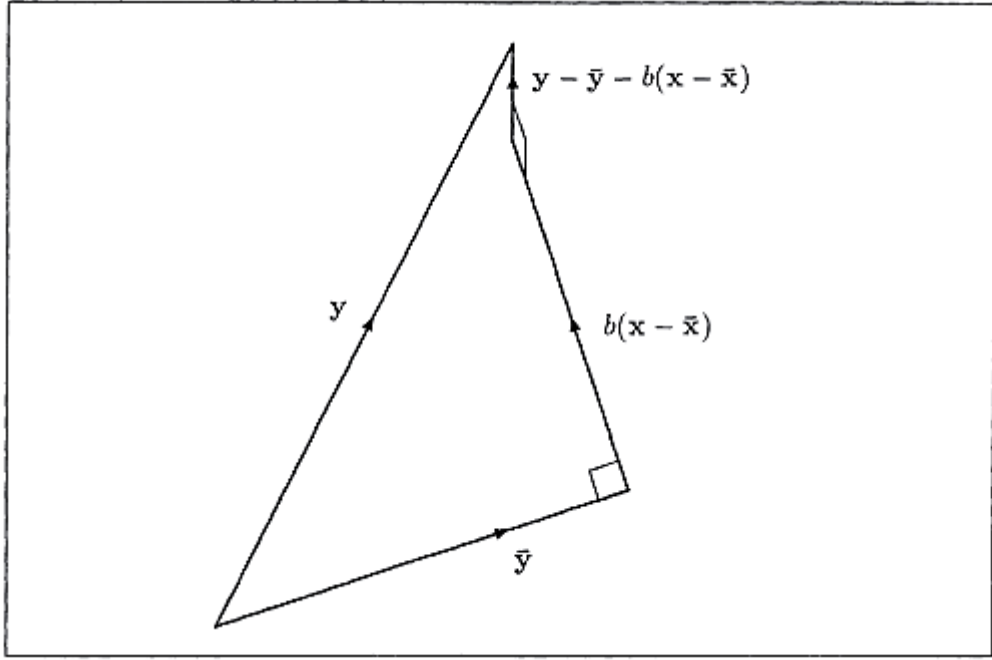
dir. Bu,

$$F = \frac{b^2 \|\mathbf{x} - \bar{\mathbf{x}}\|^2}{\|\mathbf{y} - \bar{\mathbf{y}} - b(\mathbf{x} - \bar{\mathbf{x}})\|^2 / (n-2)} = \frac{b^2 \|\mathbf{x} - \bar{\mathbf{x}}\|^2}{s^2} = \frac{[\mathbf{y}(\mathbf{x} - \bar{\mathbf{x}})]^2}{s^2 \|\mathbf{x} - \bar{\mathbf{x}}\|^2} \quad (7.50)$$

test istatistiğine götürür. Eğer $\beta = 0$ ise, bu test istatistiği bir $F_{1,(n-2)}$ dağılımını izler, bu nedenle, eğer gözlenen F değeri bu dağılımın 95 yüzdeliğini aşarsa $\beta = 0$ olduğu hipotezini reddederiz. $\beta = 0$ olup olmadığının eşdeğer t testi

$$t = \frac{b \|\mathbf{x} - \bar{\mathbf{x}}\|}{s} = \frac{\mathbf{y}(\mathbf{x} - \bar{\mathbf{x}})}{s \|\mathbf{x} - \bar{\mathbf{x}}\|} \quad (7.51)$$

dir. Eğer $\beta = 0$ ise, bu test istatistiği bir $t_{(n-2)}$ dağılımını izler.



Şekil 7.15. Basit regresyon durumu için gözlem vektörünün ortogonal ayrışımı

$\beta = 0$ olup olmadığının başka bir eşdeğer t testi, $\beta = 0$ hipotezi altında, onun dağılımının tablolanmış yüzdeleri ile r korelasyon katsayısını karşılaştırmak suretiyle verilir. Burada θ Şekil 7.13'te gösterildiği gibi, "düzeltilmiş" \mathbf{y} ve \mathbf{x} vektörleri arasındaki açı olmak üzere,

$$r = \cos \theta = \frac{\mathbf{x} - \bar{\mathbf{x}} \cdot \mathbf{y} - \bar{\mathbf{y}}}{\|\mathbf{x} - \bar{\mathbf{x}}\| \cdot \|\mathbf{y} - \bar{\mathbf{y}}\|} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (7.52)$$

dir.

8. SONUÇ ve ÖNERİLER

Yedi bölümden oluşan bu tezde geometrinin istatistikte kullanımına dair yöntemler verilmiştir. Konunun anlaşılmasındaki bazı güçlükler, geometrik yorum ve gösterimlerle ortadan kaldırılmaya çalışıldı. Özellikle verilen örneklerle geometrik yorumun nasıl kullanılabilceği gösterildi. Bunun dışında regresyon analizinde yine geometrik yorumlar ve şekillerle açıklanmaya çalışıldı. Bu tezde açıklanmaya çalışılan konular ve sunulan geometrik yorumlar istatistikte birçok alanda karşılaşılan zorlukları yeneceği kanaatindeyim.



KAYNAKLAR

- Box, J. F., Fisher, R.A. 1978. The life of a Scientist, New York: Wiley.
- Box, G.E.P., Hunter, W.G., Hunter, J.S. 1978. Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building. New York: Wiley.
- Bryant, P. 1984. Geometry, statistics, probability: Variations on a common theme. The American Statistician, 38: 38-48.
- Bartlett, M.S. 1933. The Vector Representation of a Sample. Proceedings of the Cambridge Philosophical Society, 30: 327-340.
- Chance, W.A. 1986. A geometric derivation of the distribution of the correlation coefficient $|r|$ when $\rho = 0$. The American Mathematical Monthly, 93(2) : 94-98
- Char, B.W., Geddes, K.O., Gonnet, G.H., Leong, B.L., Monagan, M.B., Watt, S.M. 1991. Maple V Language Reference Manual. New York
- Corsten, L.C.A. 1958. Vectors, a tool in statistical regression theory. Mededelingen van de Landbouwhogeschool te Wageningen, Nederland 58: 1-92.
- Demirhan, H., Parlak, Y. 2013. Çok etkenli deneysel yöntemlere geometrik yaklaşımlar ve bir uygulama. İstatistik Araştırma Sempozyumu, DIE, 10-12 Aralık 2013, Ankara.
- Durbin, J., Kendall, M.G. 1951. The Geometry of Estimation, Biometrika, 38: 507-521.
- Herr, D.G., 1980. On the history of the use of geometry in the general linear model. The American Statistician, 34: 43-47.
- Fisher, R.A. 1915. Frequency Distribution of The Value of Correlation coefficient in Samples from An Indefinitely Large Population. Biometrika, 10: 507-521
- Freund, J.E. 2004. Modern Elementary Statistics. New Jersey, 576.
- Margoils, M.S. 1979. Perpendicular projections and elementary statistics. The American Statistician, 33:131-135.
- Mirasyedi, N. 2009. İleri İstatistik Projeleri 1: İstatistiğe Geometrik Yaklaşım http://yunus.hacettepe.edu.tr/~haydarde/ist491_mirasyedi_2009.pdf(Erişim Tarihi: 23.05.2016)
- Little, T.M., Hills, F.J. 1978. Agricultural Experimentation: Design and Analysis. New York: Wiley.
- Pearson, E.S., Hartley, H.O. 1996. Biometrika Tables for Statisticians. Cambridge University Press Biometrika Trustees, Cambridge
- Saville, D.J., Wood, G.R. 1983. Statistical Analysis of the Economics of Experimental Treatments. New Zealand Journal of Experimental Agriculture, 11: 1-3.
- Saville, D.J., Wood, G.R. 1984. Outside Row Effects in Cereals Seeding Rate Trials. New Zealand Journal of Experimental Agriculture, 12: 197-202.

- Saville, D.J., Wood G.R. 1986. A Method for Teaching Statistics Using N-Dimensional Geometry. *The American Statistician*, 40: 205-214.
- Saville, D.J., Wood, G.R. 1990. Multiple Comparison Procedures – The Practical Solution, *The American Statistician*, 44: 174-180.
- Saville, D.J., Wood, G.R. 1996. *Statistical Methods : A Geometric Premier*, New York, 280
- Saville, D.J., Wood, G.R. 1997. *Statistical Methods: The Geometric Approach*, New York, 561
- Saville, D.J., Wood, G.R. 2002. A new angle on the t-test. *The Statistician*, 51(1): 99-104.
- Saville, D.J., Wood, G.R. 2005. The ubiquitous angle. *Journal of the Royal Statistical Society*, 168(1): 95-107.
- Sözer, T. 2006. *Doğrusal Cebire Giriş*. Hacettepe Üniversitesi Yayınları, Ankara, 248.

EK LİSTESİ

EK 1.

Serbestlik Dereceleri (q)	$F_{1,q}$ nun yüzdeleri		t_q ve r'nin yüzdeleri		θ' nın Yüzdeleri	
	$F_{1,q}$		t_q	r	θ	
1	90	40 95	6.31	0.988	5	0.15708
	95	161 97.5	12.71	0.997	2.5	0.07854
	99	4052 99.5	63.66	1.000	0.5	0.01571
2	90	8.5 95	2.920	0.900	5	0.45103
	95	18.5 97.5	4.303	0.950	2.5	0.31756
	99	98.5 99.5	9.925	0.990	0.5	0.14154
3	90	5.54 95	2.920	0.805	5	0.63447
	95	10.13 97.5	4.303	0.878	2.5	0.49842
	99	34.12 99.5	9.925	0.959	0.5	0.28828
4	90	4.54 95	2.353	0.729	5	0.75350
	95	7.71 97.5	3.182	0.811	2.5	0.62425
	99	21.20 99.5	5.481	0.917	0.5	0.40980
5	90	4.06 95	2.015	0.669	5	0.83734
	95	6.61 97.5	2.571	0.754	2.5	0.71592
	99	16.26 99.5	4.032	0.875	0.5	0.506634
6	90	3.78 95	1.943	0.621	5	0.90015
	95	5.99 97.5	2.447	0.707	2.5	0.78592
	99	13.75 99.5	3.707	0.834	0.5	0.58386
7	90	3.59 90	1.860	0.582	5	0.94936
	95	5.59 97.5	2.306	0.666	2.5	0.84145
	99	12.25 99.5	3.355	0.798	0.5	0.64736
8	90	3.46 95	1.860	0.549	5	0.98920
	95	5.32 97.5	2.306	0.632	2.5	0.88680
	99	11.26 99.5	3.355	0.765	0.5	0.70039
9	90	3.36 95	1.833	0.521	5	1.02230
	95	5.12 97.5	2.262	0.602	2.5	0.92471
	99	10.56 99.5	3.250	0.735	0.5	0.74544
10	90	3.29 95	1.812	0.497	5	1.05035
	95	4.96 97.5	2.228	0.576	2.5	0.95699
	99	10.04 99.5	3.169	0.708	0.5	0.78429

Serbestlik Dereceleri (q)	$F_{1,q}$ nun yüzelikleri		t_q ve r'nin yüzelikleri			θ' nin Yüzelikleri	
	$F_{1,q}$		t_q	r		θ	
11	90	3.23	95	1.796	0.476	5	1.07452
	95	4.84	97.5	2.201	0.553	2.5	0.98491
	99	9.65	99.5	3.106	0.684	0.5	0.81821
12	90	3.18	95	1.782	0.458	5	1.09561
	95	4.75	97.5	2.179	0.532	2.5	1.00935
	99	9.33	99.5	3.055	0.661	0.5	0.84815
13	90	3.14	95	1.711	0.441	5	1.11424
	95	4.67	97.5	2.160	0.514	2.5	1.03098
	99	9.07	99.5	3.012	0.641	0.5	0.87481
14	90	3.10	95	1.761	0.426	5	1.13084
	95	4.60	97.5	2.145	0.497	2.5	1.05030
	99	8.86	99.5	2.977	0.623	0.5	0.89875
15	90	3.07	95	1.753	0.412	5	1.14575
	95	4.54	97.5	2.131	0.482	2.5	1.06769
	99	8.68	99.5	2.947	0.606	0.5	0.92040
16	90	3.05	95	1.746	0.400	5	1.15925
	95	4.49	97.5	2.120	0.468	2.5	1.08346
	99	8.53	99.5	2.921	0.590	0.5	0.94009
17	90	3.03	90	1.740	0.389	5	1.17154
	95	4.45	97.5	2.110	0.456	2.5	1.09783
	99	8.40	99.5	2.898	0.575	0.5	0.95811
18	90	3.01	95	1.734	0.378	5	1.18279
	95	4.41	97.5	2.101	0.444	2.5	1.11100
	99	8.29	99.5	2.878	0.561	0.5	0.97468
19	90	2.99	95	1.729	0.369	5	1.19315
	95	4.38	97.5	2.093	0.433	2.5	1.12314
	99	8.18	99.5	2.861	0.549	0.5	0.98997
20	90	2.97	95	1.725	0.360	5	1.20281
	95	4.35	97.5	2.086	0.423	2.5	1.13436
	99	8.10	99.5	2.845	0.537	0.5	1.00416

Serbestlik Dereceleri (q)	$F_{1,q}$ nun yüzdeleri		t_q ve r'nin yüzdeleri			θ' nin Yüzdeleri	
		$F_{1,q}$		t_q	r		θ
22	90	2.95	95	1.717	0.344	5	1.21985
	95	4.30	97.5	2.174	0.404	2.5	1.15449
	99	7.95	99.5	2.819	0.515	0.5	1.02967
24	90	2.93	95	1.711	0.330	5	1.23481
	95	4.26	97.5	2.064	0.388	2.5	1.17207
	99	7.82	99.5	2.797	0.496	0.5	1.05203
26	90	2.91	95	1.706	0.317	5	1.24800
	95	4.26	97.5	2.056	0.374	2.5	1.18760
	99	7.72	99.5	2.779	0.479	0.5	1.07184
28	90	2.89	95	1.701	0.306	5	1.25975
	95	4.20	97.5	2.048	0.361	2.5	1.20145
	99	7.64	99.5	2.763	0.463	0.5	1.08954
30	90	2.88	95	1.697	0.296	5	1.27030
	95	4.17	97.5	2.042	0.349	2.5	1.21390
	99	7.56	99.5	2.750	0.449	0.5	1.10549
35	90	2.85	95	1.690	0.275	5	1.29261
	95	4.12	97.5	2.030	0.325	2.5	1.24024
	99	7.42	99.5	2.724	0.418	0.5	1.13932
40	90	2.84	90	1.684	0.257	5	1.31059
	95	4.08	97.5	2.021	0.304	2.5	1.26149
	99	7.31	99.5	2.704	0.393	0.5	1.16672
45	90	2.82	95	1.679	0.243	5	1.32548
	95	4.06	97.5	2.014	0.288	2.5	1.27911
	99	7.23	99.5	2.690	0.372	0.5	1.18948
50	90	2.81	95	1.679	0.231	5	1.33808
	95	4.03	97.5	2.009	0.273	2.5	1.29403
	99	7.17	99.5	2.678	0.354	0.5	1.20879
60	90	2.79	95	1.671	0.211	5	1.35837
	95	4.00	97.5	2.000	0.250	2.5	1.31808
	99	7.08	99.5	2.660	0.325	0.5	1.23998

Serbestlik Dereceleri (q)	$F_{1,q}$ nun yüzdeleri		t_q ve r'nin yüzdeleri			θ' nin Yüzdeleri	
	$F_{1,q}$		t_q	r		θ	
70	90	2.78	95	1.667	0.195	5	1.37414
	95	3.98	97.5	1.994	0.232	2.5	1.33678
	99	7.01	99.5	2.648	0.302	0.5	1.26428
80	90	2.77	95	1.664	0.183	5	1.38684
	95	3.96	97.5	1.990	0.217	2.5	1.35184
	99	6.96	99.5	2.639	0.283	0.5	1.28392
90	90	2.76	95	1.662	0.173	5	1.39739
	95	3.95	97.5	1.987	0.205	2.5	1.36437
	99	6.93	99.5	2.632	0.267	0.5	1.30021
100	90	2.76	95	1.660	0.164	5	1.40627
	95	3.94	97.5	1.984	0.195	2.5	1.37494
	99	6.90	99.5	2.626	0.254	0.5	1.31400
150	90	2.74	95	1.655	0.134	5	1.43684
	95	3.90	97.5	1.976	0.159	2.5	1.41084
	99	6.81	99.5	2.609	0.208	0.5	1.36091
200	90	2.73	95	1.653	0.116	5	1.45447
	95	3.89	97.5	1.972	0.138	2.5	1.43226
	99	6.76	99.5	2.601	0.181	0.5	1.38894
300	90	2.72	90	1.650	0.095	5	1.47582
	95	3.87	97.5	1.968	0.113	2.5	1.45767
	99	6.72	99.5	2.592	0.148	0.5	1.42223
400	90	2.72	95	1.649	0.082	5	1.48855
	95	3.86	97.5	1.966	0.098	2.5	1.47282
	99	6.70	99.5	2.588	0.128	0.5	1.44210

ÖZGEÇMİŞ

Adı Soyadı : Emine DEMİRCİ
Doğum Yeri : Trabzon
Doğum Tarihi : 10.09.1987
Yabancı Dili : İngilizce
E-mail : emined2011@hotmail.com
İletişim Bilgileri : Ordu Üniversitesi Fen Edebiyat Fakültesi

Öğrenim Durumu :

Derece	Bölüm/ Program	Üniversite	Yıl
Lisans	Matematik	Karadeniz Teknik Üniversitesi	2010

İş Deneyimi:

Görev	Görev Yeri	Yıl
Öğretmen	Çayeli Vakıfbank Anadolu Lisesi	2014
Öğretmen	Maçka Çok Programlı Anadolu LİSESİ	2017