BİL 2014
Çok Değişkenli Veri Analizi

Prof.Dr. Çağın Kandemir Çavaş

Dokuz Eylül Üniversitesi
Fen Fakültesi
Bilgisayar Bilimleri Bölümü

- Öğretim Üyesi
  - Prof.Dr. Çağın Kandemir Çavaş

Öğrenme Hedefleri

- Doğrusal cebir terimlerinin (determinant, özdeğer, özvektör vs.) istatistiksel anlamlarını kavrayabilme,
- Çok değişkenli tanımlayıcı istatistikleri elde edebilme (ortalama vektörü, varyans-kovaryans matrisi, korelasyon matrisi vs.),
- Hipotez Testleri ve Çok değişkenli Hipotez testlerini çözümleyebilme
- Çok değişkenli veri analiz yöntemlerini uygulayabilme,

Kaynaklar

- **Ana kaynaklar:**

  - Anderson T. W., *An Introduction To Multivariate Statistical Analysis*, Wiley-Interscience, 2003.
  - Alpar, R., *Uygulamalı Çok Değişkenli İstatistiksel Yöntemler*, Detay Yayıncılık, 2011.
  - Kazım Özdamar, Paket Programlar ile İstatistiksel Veri Analizi 2 (Çok değişkenli Analizler)
  - Grinn, L. G. and Fidell, L. S., Reading and Understanding More Multivariate Statistics, APA Books, Washington D. C., 2000.
  - Tabachnick, B. G., & Fidell, L. S., Using Multivariate Statistics, Harper Collins College Publishers, 2001

Değerlendirme

- **Arasınav** : %40
- **Final** : %60

Devamsızlık

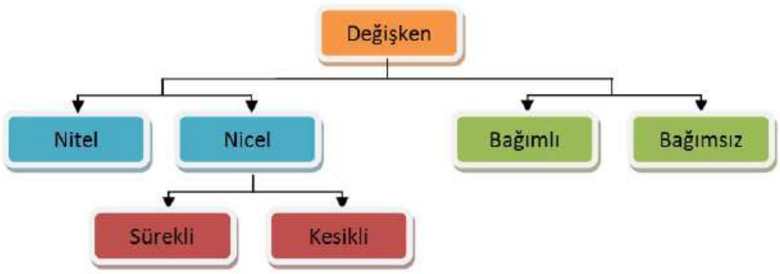
- Teorik derslerin ve öğretim üyeleri veya elemanları tarafından sınıfta yapılan uygulamaların en az %70'ine katılmış olması gerekmektedir.

İSTATİSTİKSEL TANIMLAR

Değişken

- Birimlerin farklı değerler alabilen nitelik ve niceliklerine değişken denir.
  - Meslek
  - Eğitim düzeyi
  - Gelir
  - Yaş
  - Cinsiyet…

Değişken Türleri



Değişken Türleri

- **Nitel değişken:** sayılarla ifade edilebilen değişken
- **Nicel değişken:** betimlenerek ifade edilebilen değişken
- **Sürekli değişken:** İki birim arasında sonsuz bölünebilme şansı varsa tanımlanan değişken
- **Kesikli değişken:** İki birim arasında bölünme şansı yoksa tanımlanan değişken
- **Bağımlı değişken:** Başka bir değişkene bağlı olarak değişen değişken
- **Bağımsız değişken:** Başka bir değişkene bağlı olmadan artan ya da azalan değişken

Değişken Ölçümleri

- **Sınıflama ölçme düzeyi**: Eşitlik temeline dayanır. Önemlilik düzeyi yoktur.
  - Tercih edilen marka
  - Desteklenen aday
  - Cinsiyet
  - Tedavide kullanılan ilaç türü
  - Tarımda gübre türü

- **Sıralama ölçme düzeyi**: Önem sırasına sahip ölçümü destekler. Eşitlik ve büyüklük-küçüklük esasına dayanır.
  - Başarı düzeyi
    - Çok başarılı
    - Başarılı
    - Orta
    - Başarısız
    - Çok başarısız
  - Ekonomik durum

- **Eşit Aralıklı Ölçme Düzeyi**: Sayılarla toplama, çıkarma, çarpma ve bölme yapılabilir.
  - Hava sıcaklığı
  - Başarı puanı
  - Zeka derecesi
- Ölçmenin başlangıç noktası keyfidir.

- **Oranlama ölçme düzeyi**: Eşit aralıklı düzeyde yapılan ölçme mutlak sıfır noktasına göre yapılıyorsa bu ölçme düzeyine oranlama ölçme düzeyi denir.
  - Ağırlık
  - Hız
- 100 kg bir birey 50 kg bir bireyin iki katıdır.

Veri

Yaş
27
54
35
40
52

- Tanımlanan belirli bir konuda, belirli bir amaç için belirli bir kitleye ait çeşitli özelliklerle ilgili olarak derlenmiş olan ve anlam içeren rakam, sayı, simgelerdir.

Veri Matrisi

n birimden ve p değişkenden elde edilen matrislere **veri matrisi** denir.

Birim	Yaş	Boy	*SKB
1	27	170	110
2	54	168	125
3	35	172	120
4	40	176	130
5	52	165	145

**SKB: Sistolik Kan Basıncı*

Veri Matrisi

- Tabloda yer alan verileri değişkenler sütunlarda olacak şekilde matris formunda gösterelim:

$$X_{(5x3)} = \begin{bmatrix} 27 & 170 & 110 \\ 54 & 168 & 125 \\ 35 & 172 & 120 \\ 40 & 176 & 130 \\ 52 & 165 & 145 \end{bmatrix}$$

Matrisler ve Vektörler

- **Matris:** Bir sayının düzenli sıra ve sütunlar halinde dikdörtgen biçiminde gösterilmesine matris denir.

$$A = \begin{bmatrix} 2 & 2 & 4 \\ 5 & 7 & 3 \\ 8 & 8 & 4 \end{bmatrix}$$

Matrisler ve Vektörler

- Bir matriste n tane satır p tane sütun bulunur.

$$A_{nxp} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & & a_{np} \end{bmatrix}$$

Matrisler ve Vektörler

- **Vektör:** n birimin bir özelliğini (değişken) gösteren sayıların bir sütun ya da sıra halinde gösterilmesine vektör denir.

Matrisler ve Vektörler

- Bir matriste sayılar n sıra ve 1 sütunda gösterilir ise bu matrise **sütun matris** ve ya **sütun vektör** denir.

$$a = \begin{bmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_{n-1} \\ a_n \end{bmatrix}$$

- Bir matriste sayılar bir sıra ve p sütunda gösterilir ise bu matrise **satır matris** ya da **satır vektör** denir.

$$a^T = \begin{bmatrix} a_1 & a_2 & . & . & . & a_{n-1} & a_n \end{bmatrix}$$

Matris Çeşitleri

- Bir matrisin herhangi bir sıra ya da sütununda yer alan sayısal değere **matrisin elemanı** adı verilir ve a_{ij} ile gösterilir.
- Eğer i=j ise a_{ij} değeri bize **köşegen (diagonal)** elemanlarını verir.

Matris Çeşitleri

- **Kare matris:** Eğer bir matriste n=p ise yani satır sayısı=sütun sayısı ise o matrise kare matris denir. Tüm i=j olan elemanlarına ise esas köşegen elemanları adı verilir.
- **Sıfır matris:** Eğer tüm i,j için $a_{ij} = 0$ ise a matrisine 0 matrisi denir.

Matris Çeşitleri

- **Köşegen (diagonal) matris**: Esas köşegen elemanlarının skaler değerler köşegen dışı olan elemanların ise 0 olduğu matristir.

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

Matris Çeşitleri

- **Birim Matris (I):** Bir kare matrisin tüm i=j olan elemanlarının 1 geri kalan elemanlarının 0 olması durumudur.

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Matris Çeşitleri

- **Skaler Matris:** Birim matrisin herhangi bir skaler ile çarpılması sonucu elde edilen matris.

$$A = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

Matris Çeşitleri

- **Devrik (Transpoz) Matris**: Bir A matrisinin satır ve sütunlarının yer değiştirmesi işlemidir. A′ ile gösterilir.

Örnek:

$$A = \begin{bmatrix} 2 & 3 & 8 \\ 4 & 8 & 7 \end{bmatrix} \qquad A' = \begin{bmatrix} 2 & 4 \\ 3 & 8 \\ 8 & 7 \end{bmatrix}$$

Matris Çeşitleri

- Devrik matrisin özellikleri:

  I′=I

  (A′)′=A

  (A+B)′=A′+B′

  (AB)′=B′A′

  (kA)′=kA′

  (ABC)′=C′B′A′

Matris Çeşitleri

- **Simetrik Matris:** Bir A kare matrisinde A′=A ise A matrisine simetrik matris denir.

$$A=\begin{bmatrix}7 & 3\\ 3 & 9\end{bmatrix} \qquad A'=\begin{bmatrix}7 & 3\\ 3 & 9\end{bmatrix}$$

$$B=\begin{bmatrix}4 & 8 & 7\\ 8 & 6 & 2\\ 7 & 2 & 1\end{bmatrix} \qquad B'=\begin{bmatrix}4 & 8 & 7\\ 8 & 6 & 2\\ 7 & 2 & 1\end{bmatrix}$$

Matris Çeşitleri

- **Çarpık Simetrik Matris**: A=-A′ ise A matrisine çarpık simetrik matris denir.

$$A=\begin{bmatrix}0 & 1 & 2\\ -1 & 0 & -3\\ -2 & 3 & 0\end{bmatrix} \qquad -A'=\begin{bmatrix}0 & 1 & 2\\ -1 & 0 & -3\\ -2 & 3 & 0\end{bmatrix}$$

Matris Çeşitleri

- **Matrislerin Eşitliği:** Bir matrisin her bir elemanı, diğer bir matrisin karşı gelen elemanına eşit ise iki matris eşittir.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

$$\left.\begin{array}{l} a_{11} = b_{11} \\ a_{12} = b_{12} \\ . \\ . \\ . \\ a_{33} = b_{33} \end{array}\right\} \Longrightarrow \text{A=B}$$

Matris Çeşitleri

- **Bir matrisin izi (Trace)**: Bir kare matrisin izi köşegen elemanlarının toplamına eşittir.

$$tr(A) = \sum_{i=1}^{n} a_{ii}$$

$$A = \begin{bmatrix} 6 & 8 & -1 \\ 5 & 2 & -6 \\ 3 & 3 & 1 \end{bmatrix} \qquad tr(A) = 6 + 2 + 1 = 9$$

Matrislerde Aritmetik İşlemler *-Toplama ve Çıkarma*

- Matrislerde toplama ve çıkarma işleminin yapılabilmesi için matris boyutlarının aynı olması gerekir.

$$A = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{21} & b_{31} \\ b_{12} & b_{22} & b_{32} \\ b_{13} & b_{23} & b_{33} \end{bmatrix}$$

Matrislerde Aritmetik İşlemler *-Toplama ve Çıkarma*

- Sırası ile A+B ve A-B matrislerini yazalım.

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{21} + b_{21} & a_{31} + b_{31} \\ a_{12} + b_{12} & a_{22} + b_{22} & a_{32} + b_{31} \\ a_{13} + b_{13} & a_{23} + b_{23} & a_{33} + b_{33} \end{bmatrix}$$

$$A - B = \begin{bmatrix} a_{11} - b_{11} & a_{21} - b_{21} & a_{31} - b_{31} \\ a_{12} - b_{12} & a_{22} - b_{22} & a_{32} - b_{31} \\ a_{13} - b_{13} & a_{23} - b_{23} & a_{33} - b_{33} \end{bmatrix}$$

Matrislerde Aritmetik İşlemler
-*Toplama ve Çıkarma*

- Örnek 1: Aşağıda verilen A,B ve C matrisleri veriliyor. Sırası ile;

  A+B+C, A-B-C, A+B-C matrislerini bulunuz.

$$A = \begin{bmatrix} 2 & 1 \\ 4 & 5 \end{bmatrix} \quad B = \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix} \quad C = \begin{bmatrix} 2 & -5 \\ 0 & 4 \end{bmatrix}$$

Matrislerde Aritmetik İşlemler
-*Çarpma*

- Matrisin Skaler bir sabit ile çarpılması:

$$kA = \begin{bmatrix} ka_{11} & ka_{21} & ka_{31} \\ ka_{12} & ka_{22} & ka_{23} \\ ka_{13} & ka_{23} & ka_{33} \end{bmatrix}$$

Matrislerde Aritmetik İşlemler *-Çarpma*

- A matrisinin k=-5 ile çarpımından oluşan yeni matrisi bulunuz.

$$A = \begin{bmatrix} 0 & 1 \\ 7 & -2 \end{bmatrix}$$

- A matrisinin k=-0.5 ile çarpımından oluşan yeni matrisi bulunuz.

Matrislerde Aritmetik İşlemler *-Çarpma*

- **İki matrisin çarpımı**: İki matrisin çarpılabilmesi için soldaki matrisin sütun sayısının sağdaki matrisin satır sayısına eşit olması gerekir. Örneğin;

$$A_{23} B_{33} = C_{23}$$

Matrislerde Aritmetik İşlemler -*Çarpma*

- A ve B matrislerinin çarpımı aşağıdaki şekilde hesaplanır.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$AB = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{bmatrix}$$

Determinant

- Matrislerin tersini bulmak için öncelikle determinantını bulmak gereklidir.
- 2x2 boyutlu bir matrisin determinantı:

$$\det(A) = \Delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

Determinant

- Minörler, Kofaktörler, Adjoint Matris:

- A matrisinin i. satır ve j. sütunu çıkartıldığında geriye kalan matrisin determinantına a_{ij}'nin **minörü** denir ve $|M_{ij}|$ şeklinde gösterilir.

Determinant

- $\alpha_{ij} = (-1)^{i+j}|M_{ij}|$ biçimindeki ifadeye ise a_{ij}'nin kofaktörü denir.

- A matrisi

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Determinant

- 1. satır için minörler ve kofaktörleri bulalım:

$$|M_{11}| = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \quad |M_{12}| = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} \quad |M_{13}| = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$\alpha_{11} = (-1)^{1+1}|M_{11}|$$

$$\alpha_{12} = (-1)^{1+2}|M_{12}|$$

$$\alpha_{13} = (-1)^{1+3}|M_{13}|$$

Determinant

- Bir matrisin determinant değeri A'nın herhangi bir satır yada sütunundaki her elemanın kendi kendi kofaktörü ile çarpımlarının toplamına eşittir.

$$|A| = a_{11}\alpha_{11} + a_{12}\alpha_{12} + a_{13}\alpha_{13}$$

Determinant

- A matrisinin elemanlarını kofaktörler cinsinden yazalım.
- Bu matrisin transpozu adjoint matrisini verir.

$$A_K = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{13} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix} \qquad A'_K = Adj(A) = \begin{bmatrix} \alpha_{11} & \alpha_{21} & \alpha_{31} \\ \alpha_{12} & \alpha_{22} & \alpha_{32} \\ \alpha_{13} & \alpha_{23} & \alpha_{33} \end{bmatrix}$$

Bir Matrisin Tersi (Inverse)

- Matrisin tersi A^{-1} şeklinde gösterilir.
- Bir matrisin kendisi ile tersinin çarpımı birim matrise eşittir.

$$A^{-1}A = I$$

Bir Matrisin Tersi (Inverse)

- Matris denklem sistemlerini çözmek için matrislerin tersini bulmaya ihtiyaç vardır. Aşağıda X vektörünü bulmak için A matrisinin tersine ihtiyaç vardır:

$$AX = B \qquad X = A^{-1}B$$

Bir Matrisin Tersi (Inverse)

- Bir A matrisinin tersi aşağıdaki şekilde bulunur:

$$A^{-1} = \frac{1}{\det(A)} Adj(A)$$

- Not: Eğer bir matrisin **determinant değeri 0** ise **tersi hesaplanamaz.**

Bir Matrisin Rankı

- Determinantı 0 olan matrise singular yani tekil matris denir.
- Eğer matris **kare matris değilse** ve ya **determinant değeri 0 a eşit** ise **tersi hesaplanamaz**.
- Determinantı sıfırdan farklı olmak kaydıyla bir matrisin sıra ve sütun sayılarından **küçük** olanına **matrisin rankı** denir

Bir Matrisin Rankı

- Örneğin aşağıda verilen A matrisini inceleyelim. A matrisinin alt matrislerini yazmaya çalışalım.

$$A = \begin{bmatrix} 6 & 4 \\ 3 & 2 \\ -1 & 4 \end{bmatrix} \qquad \begin{bmatrix} 6 & 4 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ -1 & 4 \end{bmatrix}$$

Bir Matrisin Rankı

- İlk matrisin determinantı 0'dır. Geriye kalan matrislerin boyutu 2x2 olduğundan yani en büyük derece 2 olduğundan matrisin rankı 2'dir ve r(A)=2 şeklinde gösterilir.

Uygulamalar

$$A = \begin{bmatrix} -1 & 3 \\ 4 & -2 \\ 5 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} -3 & 2 \\ -4 & 1 \end{bmatrix}$$

AB yi bulunuz.

Çözüm:

$$AB = \begin{bmatrix} (-1)(-3)+(3)(-4) & (-1)(2)+(3)(1) \\ (4)(-3)+(-2)(-4) & (4)(2)+(-2)(1) \\ (5)(-3)+(0)(-4) & (5)(2)+(0)(1) \end{bmatrix}$$

$$= \begin{bmatrix} -9 & 1 \\ -4 & 6 \\ -15 & 10 \end{bmatrix}$$

Uygulamalar

1. A ve B kare matrisleri veriliyor. AB ve BA matrislerinin birbirine eşit olmadığını doğrulayınız.

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$AB = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \qquad BA = \begin{bmatrix} 3 & 0 \\ 1 & 1 \end{bmatrix}$$

$$AB \neq BA$$

Uygulamalar

2. (A′)′=A olduğunu verilen A matrisi ile doğrulayınız.

$$A = \begin{bmatrix} 2 & -5 \\ 0 & 4 \end{bmatrix}$$

$$A' = \begin{bmatrix} 2 & 0 \\ -5 & 4 \end{bmatrix}$$

$$(A')' = \begin{bmatrix} 2 & -5 \\ 0 & 4 \end{bmatrix}$$

Uygulamalar

3. X vektörünün X’ ile çarpımının, vektör elemanlarının kareleri toplamına eşit olduğunu doğrulayınız.

$$X = \begin{bmatrix} 2 \\ 5 \\ 9 \end{bmatrix}$$

$$X' = \begin{bmatrix} 2 & 5 & 9 \end{bmatrix}$$

$$X'X = 2^2 + 5^2 + 9^2$$
$$= 110$$

Uygulamalar

4. Aşağıda U ve A matrisi verilmiştir. U’AU çarpımını hesaplayınız.

$$U = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} \qquad A = \begin{bmatrix} 5 & 2 & -1 \\ -3 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

$$\begin{bmatrix} 3 & -1 & 2 \end{bmatrix} \begin{bmatrix} 5 & 2 & -1 \\ -3 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} = 56$$

Uygulamalar

5.Aşağıda A matrisi verilmiştir. A'A ve AA' matrislerinin simetrik olduklarını doğrulayınız.

$$A = \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix}$$

$$A'A = \begin{bmatrix} 29 & 13 \\ 13 & 17 \end{bmatrix}$$

$$AA' = \begin{bmatrix} 26 & 14 \\ 14 & 20 \end{bmatrix}$$

Uygulamalar

6.A matrisinin aynı boyutlu birim matris ile çarpımının kendisine eşit olduğunu doğrulayınız.

$$A = \begin{bmatrix} 5 & 2 & 1 \\ 3 & 4 & 5 \\ 1 & 0 & 3 \end{bmatrix}$$

$$AI = \begin{bmatrix} 5 & 2 & 1 \\ 3 & 4 & 5 \\ 1 & 0 & 3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$AI = A$$

Uygulamalar

7. A ve B kare matrisleri verilmiştir.
   a. tr(A+B)=tr(A)+tr(B) olduğunu doğrulayınız.
   b. tr(AB)=tr(BA) olduğunu doğrulayınız.

$$A=\begin{bmatrix}5 & 1\\ 2 & 4\end{bmatrix} \quad B=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$$

$$A+B=\begin{bmatrix}6 & 2\\ 2 & 5\end{bmatrix} \quad AB=\begin{bmatrix}5 & 6\\ 2 & 6\end{bmatrix} \quad BA=\begin{bmatrix}7 & 5\\ 2 & 4\end{bmatrix}$$

Uygulamalar

8. Aşağıda verilen A matrisinin tersinin olmadığını gösteriniz.

$$A=\begin{bmatrix}1 & 2 & 3\\ 2 & 3 & 1\\ 1 & 2 & 3\end{bmatrix}$$

Uygulamalar

9. Aşağıda A ve B matrisleri verilmiştir. Inv(A)B matrisini hesaplayınız.

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 7 \\ 9 & 8 & 6 \end{bmatrix}$$

$$B = \begin{bmatrix} 2 & 1 & 7 \\ 8 & 6 & 9 \\ 5 & 4 & 3 \end{bmatrix}$$

Örneklem ve Örnekleme

- **Örneklem:** Araştırılmak istenen bir olayla ilgili kütleden, belli kurallara göre seçilmiş, kütleyi temsil ettiği varsayılan küçük bir küme örneklem olarak adlandırılır.
- **Örnekleme:** Anakütleden örnek seçme işlemine denir. Örneklemin anakütleyi temsil edebilir nitelikte olması gerekir.

Örneklem için kurallar

Anakütleyi temsil yeteneğine sahip bir örneklemin temel özellikleri şunlardır.

- Örneklemin büyüklüğü (hacmi, miktarı) yeterli olmalıdır.
- Örneklem anakütledeki dağılıma çeşit ve oran yönünden benzer olmalıdır.
- Anakütledeki bütün birimlerin örneğe girme şansı eşit olmalıdır.

İstatistiksel Yöntemler

- Betimleyici (Descriptive) Yöntemler
  - ⇒ Verideki herhangi bir dağılımı bir ya da birden çok katsayıda anlatabilmek
    - Örn: sınıftaki not ortalaması
- Açıklayıcı (Explanatory) Yöntemler
  - ⇒ Bir veri setinde olası ilişkilerin belirtilmesi
    - Örn: Günlük harcama miktarı ile gelir arasındaki ilişki

Betimleyici Yöntemler

- Amaç: Eldeki dağılımı en iyi şekilde temsil etmek
- Araçlar:
  - *Ortalama*
  - *Medyan*
  - *Mod*
  - *Varyans*

Kitle varyansı

- Tüm kitle verilerin ortalama değerden sapma karelerinin ortalaması

- Kitle varyansı:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

μ = kitle ortalaması

N = kitle büyüklüğü

x_i = x değişkenin i^{inci} değeri

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

Kitle standart sapması

Örneklem varyansı

Tüm örneklem verilerin ortalama değerden sapma karelerinin ortalaması

- Örneklem varyansı: $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

$\bar{X}$ = aritmetik ortalama

n = örneklem büyüklüğü

X_i = X değişkeninin i[inci] değeri

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Örneklem standart sapması

Örneklem Verisi(x_i) :

10	12	14	15	17	18	18	24

n = 8 Ortalama = $\bar{x}$ = 16

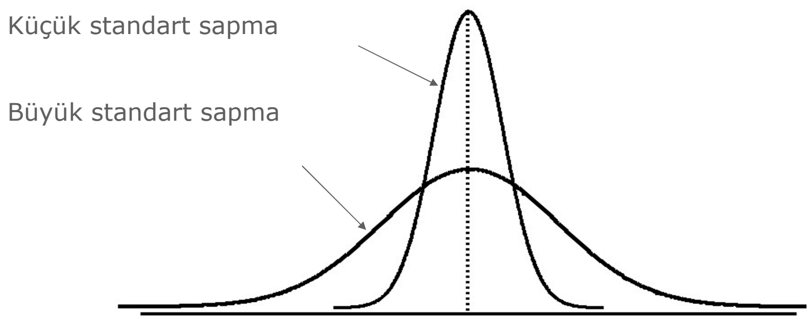
$$s = \sqrt{\frac{(10 - \bar{X})^2 + (12 - \bar{x})^2 + (14 - \bar{x})^2 + \cdots + (24 - \bar{x})^2}{n - 1}}$$

$$= \sqrt{\frac{(10 - 16)^2 + (12 - 16)^2 + (14 - 16)^2 + \cdots + (24 - 16)^2}{8 - 1}}$$

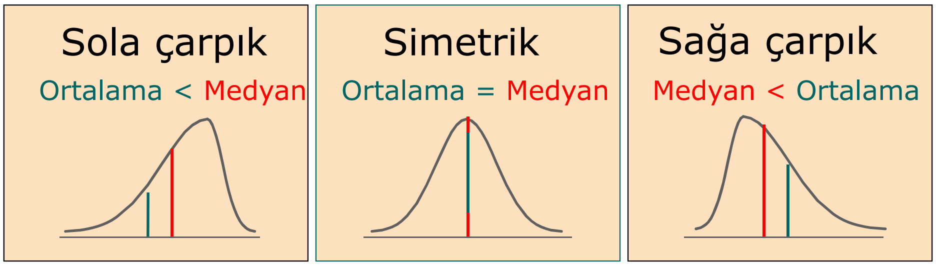
$$= \sqrt{\frac{130}{7}} = \boxed{4.3095}$$

Verilerin ortalama etrafındaki saçılımı

Değişim gösterimi



Değişkenlerin dağılım şekli



Betimleyici istatistiğin önemi

Dağılım: 6,6,6 Ortalama: 6 Medyan: 6 Mod: 6 Std. Sapma: 0	Dağılım: 0,6,12 Ortalama: 6 Medyan: 6 Mod: 6 Std: Sapma: 6

Açıklayıcı Analizler

- Amaç: Bir veri setinde olası ilişkileri ortaya çıkarmak ya da hipotezleri test etmek
- Örnek : Y= f(x)
  - ie: İnternet kullanımı= f(cinsiyet)
  - ie: Yaşam biçimi= f(gelir)
  - ie: Tüketim= f(yaşam biçimi)

Daha gelişmiş analiz

- **Covariance (kovaryans):**
  - İki değişken arasındaki doğrusal ilişkinin gücünü ölçer.
- İki değişken arasındaki kovaryans değeri:

Cov(x,y) > 0	x ve y aynı yönde hareket eder.
Cov(x,y) < 0	x ve y ters yönde hareket eder.
Cov(x,y) = 0	x ve y birbirinden bağımsızdır.

- **Kitle kovaryansı**

$$\mathrm{Cov}(x,y) = \sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N}$$

- **Örneklem kovaryansı**

$$\mathrm{Cov}(x,y) = s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Korelasyon Katsayısı

İki değişken arasındaki doğrusal ilişkinin bağılkuvvetini ölçer.

- **Kitle korelasyon katsayısı:**

$$\rho = \frac{Cov(x,y)}{\sigma_X \sigma_Y}$$

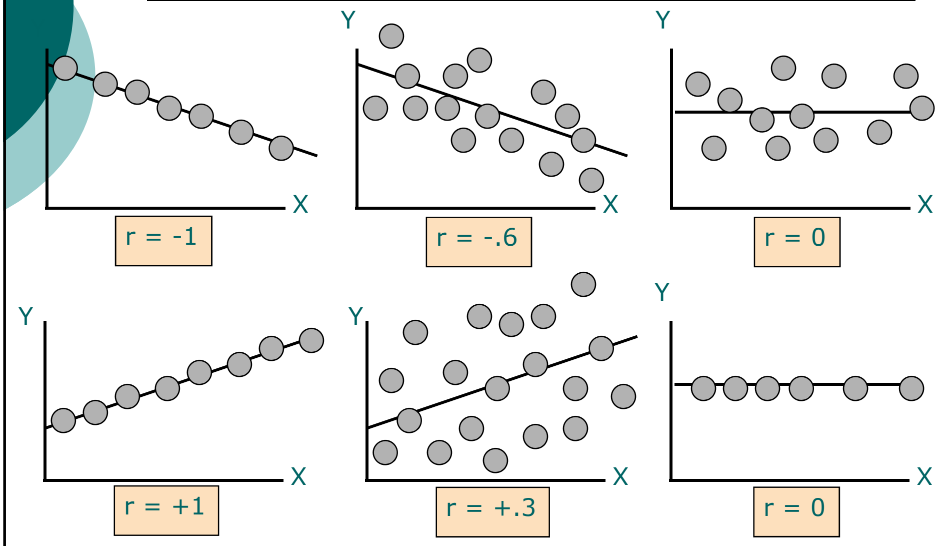
- **Örneklem korelasyon katsayısı:**

$$r = \frac{Cov(x,y)}{s_X s_Y}$$

Korelasyon katsayısının(r) özellikleri

- Birimsiz
- −1 ve 1 arasında değerler alır.
- −1 değerine yakınsa, güçlü negatif ilişki var.
- 1 değerine yakınsa, güçlü pozitif ilişki var.
- 0 değerine yakınsa, zayıf doğrusal ilişki var.

Korelasyon Katsayılarına göre Verilerin Serpme Grafikleri



Çok Değişkenli Veri Analizi

Prof.Dr. Çağın Kandemir Çavaş
DEÜ Fen Fakültesi
Bilgisayar Bilimleri Bölümü

Hipotez Testleri

- Gözlem ya da deneme sonucu elde edilmiş sonuçların, rastlantıya bağlı olup olmadığının incelenmesinde kullanılan istatistiksel yöntemlere **hipotez testleri** denir.

Hipotez Testlerinin İçeriği

- Hipotez testi bir kitle parametresinin örnekleme ait elde edilen değerle (örneklem istatistiği) karşılaştırılıp test edilmesidir.
  - Kitle ortalaması

Örnek: İzmir'deki konutlarda aylık ortalama su faturası $\mu = 100$ TL

Örnek: DEÜ Bilgisayar Bilimleri Bölümü öğrencilerinin genel not ortalaması $\mu = 3.20$

Eğer örneklem istatistiği, test edilen kitle parametresine yakın değerde ise **hipotez doğru** olarak kabul edilir.

Sıfır Hipotezi, H_0

- Test edilecek varsayım (sayı değeri) yazılır.

Örnek:Türkiye'de evlerde bulunan ortalama TV sayısı 3 olarak düşünülmektedir.($H_0 : \mu = 3$)

H_0 Hipotezi her zaman kitle parametresi olarak yazılmalıdır, örnek istatistiği olarak değil!

$H_0 : \mu = 3$ $\qquad$ $\cancel{H_0 : \bar{x} = 3}$

Sıfır Hipotezi, H_0

(devamı)

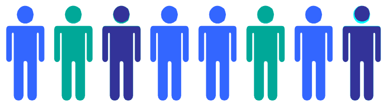
- Sıfır hipotezi doğru varsayımı ile başlanır
  - Aksi kanıtlanana kadar doğru olarak düşünülür.
- Her zaman **"="** , **"≤"** ya da "≥" işaretlerini içerir.
- Sonuçta **reddedilebilir** ya da **reddedilemez**.

Alternatif Hipotezi, H_1

- Sıfır hipotezinin tersidir.
  - **Örnek:** Türkiye'de evlerde bulunan ortalama TV sayısı 3'e eşit değildir. (H_1: $\mu \neq 3$)
- Hiçbir zaman **"="** işaretini **içermez**.
- Genellikle araştırmacı tarafından desteklenen bir hipotezdir.
- Sonuçta **reddedilebilir** ya da **reddedilemez**

Hipotez Testi Süreci

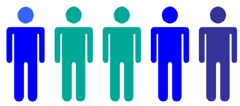
İddia:
kitlenin
Ortalama yaşı 50.
(Sıfır Hipotezi:
H_0: μ = 50)



Kitle

Şimdi örneklem seçelim

$\overline{x}$=20 ise μ = 50 hipotezi desteklenebilir mi?



Örneklem

Örneklemin ortalama yaşı 20 olsun: $\overline{x}$ = 20

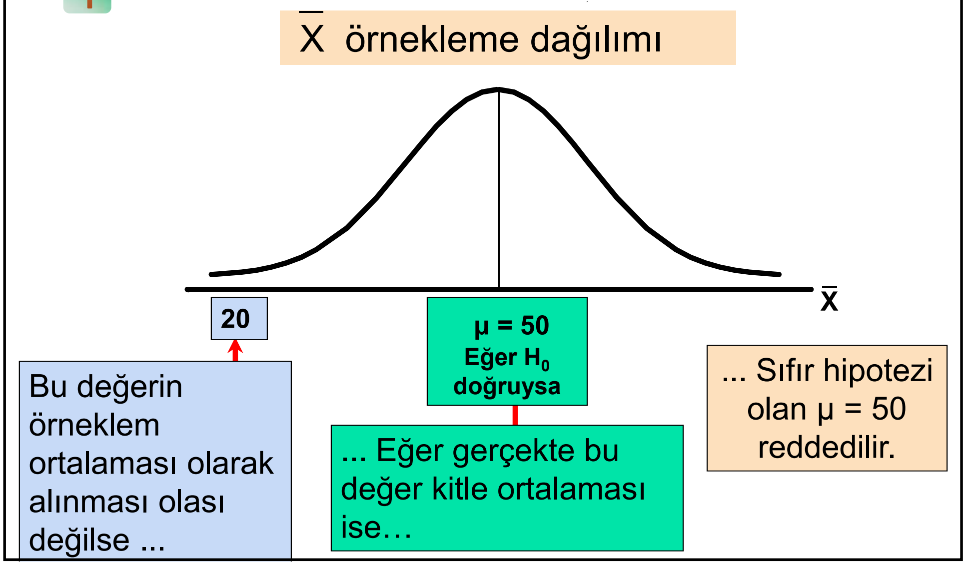
Desteklenemez ise
Sıfır Hipotezi, Ho
REDDEDİLİR

Örnek

- Bir üretici tarafından piyasaya sürülen kabloların ortalama dayanma gücü 600 kg. olsun. İmalat işlemindeki yeni bir teknikle bu gücün artırılabileceği iddia ediliyorsa bu durumda sıfır hipotez ve alternatif hipotez nasıl olmalıdır?

H_0 :μ=600 kg. (μ:ortalama)
H_1:μ>600 kg.

H_0 neden reddedilir?



Anlamlılık Seviyesi, α

- **Eğer sıfır hipotezi doğru ise, örneklem istatistiğinin olası olmayan değerlerini tanımlar.**
  - Örnekleme dağılımının red bölgesini tanımlar.
- α ile gösterilir, (anlamlılık seviyesi)
  - Genellikle 0.01, 0.05, ya da 0.10 değerlerini alır.
- Araştırmanın en başında araştırmacı tarafından seçilir. Testin kritik değerlerini verir.

Hipotez Testinin Yönü

Parametre	Testin yönü		
	İki yönlü test	Tek yönlü test	
		Sağ kuyruk	Sol kuyruk
Ortalama(μ)	H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$	H_0: $\mu = \mu_0$ H_1: $\mu > \mu_0$	H_0: $\mu = \mu_0$ H_1: $\mu < \mu_0$

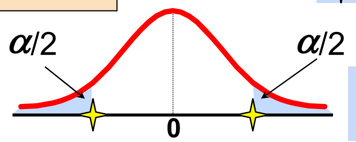
Anlamlılık Seviyesi ve Red Bölgesi

Anlamlılık seviyesi = α

kritik değerler

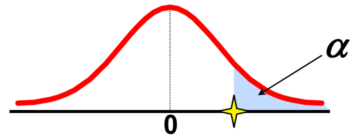
Taralı alan: red bölgesi

H_0: $\mu = 3$
H_1: $\mu \neq 3$



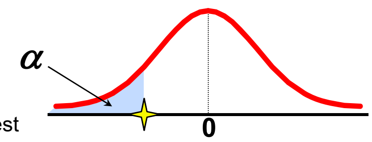
İki Yönlü Test

H_0: $\mu \leq 3$
H_1: $\mu > 3$



Üst kuyruk test

H_0: $\mu \geq 3$
H_1: $\mu < 3$



Alt kuyruk test

Karar vermedeki Hatalar

		Hipotez testi sonucunda verilen karar	
		H_0 kabul edildi	H_0 Reddedildi
Gerçek	H_0 doğru	Doğru karar (1-α) →Güven düzeyi	I. Tip Hata (α)→α hatası
	H_0 yanlış	II. Tip Hata (β)	Doğru Karar (1-β)→Testin Gücü

Karar vermedeki Hatalar

- **1.Tip Hata**
  - Doğru olan sıfır hipotezinin (H_0) reddedilmesi

1.Tip hata olasılığı α 'dır.

- Testin anlamlılık seviyesi olarak adlandırılır.
- Başlangıçta araştırmacı tarafından seçilir.

Karar vermedeki Hatalar

(devamı)

- **2.Tip Hata**
  - Yanlış olan sıfır hipotezini reddetmeme

2.Tip hata olasılığı β'dır.

Sonuçlar ve Olasılıklar

Hipotez Testi olası sonuçları

	Gerçekte olan	
Karar	H_0 Doğru	H_0 Yanlış
H_0 Kabul	**Doğru Karar** **(1 - α)**	**2.Tip Hata** **(β)**
H_0 red	**1.Tip Hata** **(α)**	**Doğru Karar** **(1 - β)**

Sonuç **(olasılık)**

(1 - β) : Testin Gücü

Kitlenin ortalaması için Hipotez Testleri

μ için Hipotez Testleri

- **σ Biliniyorsa**
- **σ Bilinmiyorsa**

Normal dağılan bir kitlenin ortalama testi (σ Biliyorsa)

- Örneklem değeri $(\bar{x})$ z değerine dönüştürülür

μ için Hipotez Testleri

- **σ Biliniyorsa**
- **σ Bilinmiyorsa**

Test aşağıdaki gibi olsun

$$H_0: \mu = \mu_0$$
$$H_1: \mu > \mu_0$$

(Kitle normal dağılıyor ve

n>30 veya n<30 olabilir)

Karar kuralı:

$$\text{Red } H_0 \text{ eger} \quad z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_\alpha$$

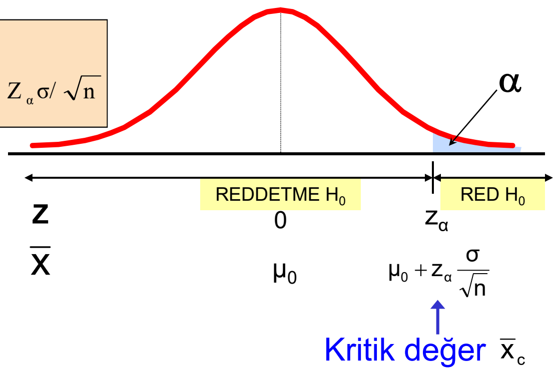
Karar Kuralı

$H_0: \mu = \mu_0$
$H_1: \mu > \mu_0$

RED H_0 eger $\quad z = \dfrac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_\alpha$

Alternatif kural:

Red H_0 eger $\quad \bar{x} > \mu_0 + Z_\alpha \sigma / \sqrt{n}$



Z-tablo değeri

- http://www.z-table.com/

Örnek-1: Ortalama için Üst kuyruk Z Testi (σ Biliniyorsa)

Bir telefon şirketi yeni reklam kampanyası sebebiyle müşterilere ait aylık telefon faturasının artacağını düşünüyor. Şuandaki aylık ortalama fatura 52 TL. Faturanın artacağını düşünen şirketin bu iddiasını test edin. (σ = 10 olarak bilindiği varsayılıyor)

Hipotez testini oluşturun:

$H_0: \mu = 52$ ortalama aylık 52 TL

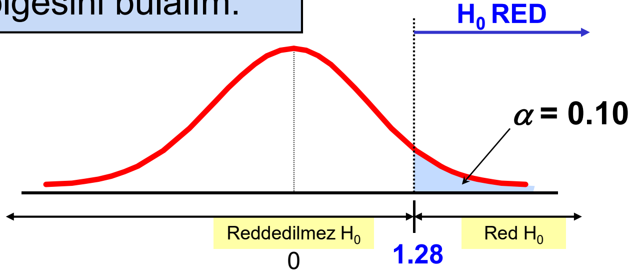
$H_1: \mu > 52$ ortalama aylık 52 TL'den fazla olacak
(şirketin bu iddiası istatistiksel olarak kanıtlanabilir mi)

Örnek-1: Red Bölgesini Bulma

(devamı)

- Bu test için α = 0.10 verilmiş olsun.

Red bölgesini bulalım:



$$\text{Red } H_0 \text{ eger } z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > 1.28$$

Örnek-1: Örneklem Sonucu

(devamı)

Örneklem istatistiği değerlerine göre test istatistiği hesaplanır.

Örnekleme ait istatistikler değerlerinin aşağıdaki gibi verildiğini varsayalım: $n = 64$, $\bar{x} = 53.1$ ($\sigma = 10$ olarak bilindiği verilmişti)

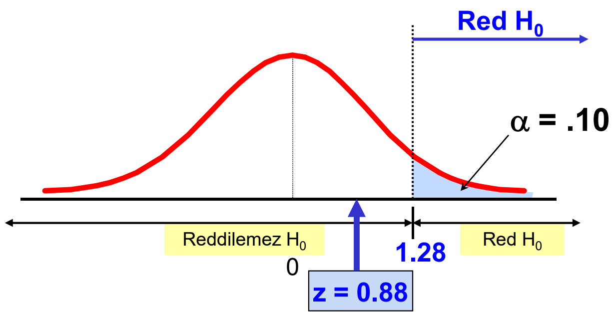
- Örneklem istatistikleri kullanarak z test istatistiğini hesaplayalım,

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{53.1 - 52}{\frac{10}{\sqrt{64}}} = 0.88$$

Örnek-1: Karar

(devamı)

Bir karara ulaşalım ve kararı yorumlayalım:



H_0 reddedilemez çünkü z = 0.88 < 1.28

yorum: Şirketin iddiasındaki gibi aylık faturanın 52 TL'nin üstünde olması beklenemez.

Tek kuyruk testleri

- Birçok durumda, alternatif hipotez belirli tek bir yöne odaklanır.

H_0: $\mu \leq 3$
H_1: $\mu > 3$

→ Bu üst-kuyruk testidir çünkü alternatif hipotez ortalama 3'ün üstündeki değerlere odaklanır

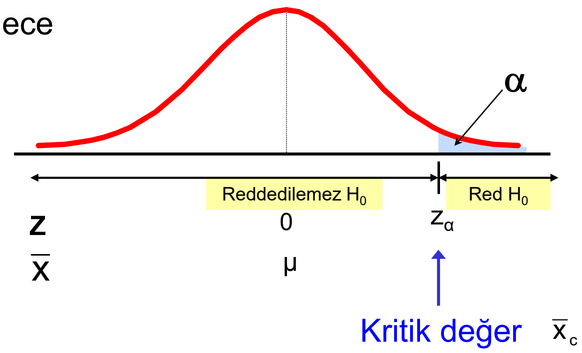
H_0: $\mu \geq 3$
H_1: $\mu < 3$

→ Bu alt-kuyruk testidir çünkü alternatif hipotez ortalama 3'ün altındaki değerlere odaklanır.

Üst-Kuyruk Testleri

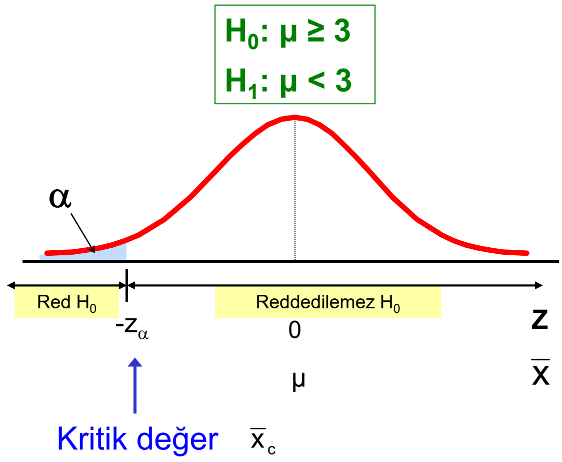
- Tek bir tane kritik değer vardır, çünkü red bölgesi sadece bir kuyrukta bulunmaktadır.

H_0: $\mu \leq 3$
H_1: $\mu > 3$



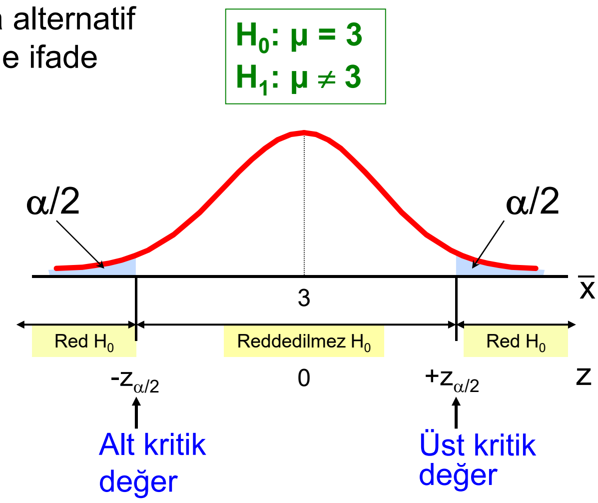
Alt-Kuyruk Test

- Tek bir tane kritik değer vardır, çünkü red bölgesi sadece bir kuyrukta bulunmaktadır.



İki Yönlü Testler

- Bazı durumlarda alternatif hipotez tek yönde ifade edilmeyebilir.
- İki tane kritik değer vardır, 2 tane red bölgesi bulunur.



Hipotez testi Örneği (σ Biliniyorsa)

Türkiye'de her ailenin sahip olduğu çocuk sayısı ortalamasının 3'ten farklı olduğu iddiasını kanıtlayınız. **(σ = 0.8 ve α = 0.05 varsayın)**

- Sıfır ve alternatif hipotezleri belirleyiniz
  - H_0: $\mu = 3$, H_1: $\mu \neq 3$ (İki-yönlü test)
- Anlamlılık seviyesi α = .05 olarak verilmiş.
- Örneklem büyüklüğünü seçiniz
  - Örneklem büyüklüğü n = 100 olarak seçilmiş olsun

Hipotez testi Örneği (σ Biliniyorsa)

(devamı)

- Uygun çözüm testini belirleyelim
  - σ biliniyorsa z testi kullanılacak
- Tablo kritik değerlerini belirleyelim
  - α = .05 için kritik z değerleri ±1.96
- Örnekleme ait istatistikleri hesapla ($\bar{x}$ ve σ biliyor) ve test istatistiğini hesapla.
  - Örnekleme ait aşağıdaki sonuçların verildiğini varsayalım

    n = 100, $\bar{x}$ = 2.84 (σ = 0.8 olarak biliyor)

Test istatistiği:

$$z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{2.84 - 3}{\frac{0.8}{\sqrt{100}}} = \frac{-.16}{.08} = -2.0$$

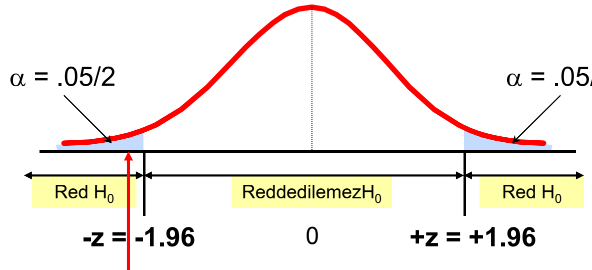
Hipotez testi Örneği (σ Biliniyorsa)

(devamı)

- Test istatistiği red bölgesinde mi?

H_0 red eğer

$z < -1.96$ veya $z > 1.96$; aksi durumda H_0 reddedilemez

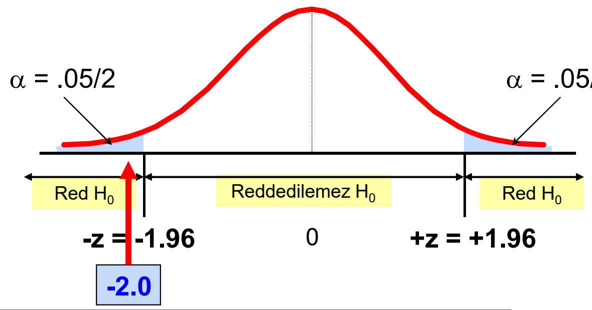


$z = -2.0 < -1.96$, olduğu için test istatistiği red bölgesindedir.

Hipotez testi Örneği (σ Biliniyorsa)

(devamı)

- Bir karara ulaşılır ve sonuç yorumlanır.



$z = -2.0 < -1.96$, olduğu için sıfır hipotezini reddederiz ve Türkiye'de her ailenin sahip olduğu çocuk sayısı ortalamasının 3'ten farklı olduğuna dair yeterli kanıt vardır denir.

Normal dağılan bir kitlenin ortalama testi (σ Bilinmiyorsa ve n<30 ise)

- Örneklem istatistiği ($\bar{x}$) t test istatistiğine dönüştürülür.

μ için hipotez testi

σ biliniyorsa | **σ Bilinmiyorsa**

Hipotez testi:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

(kitle normal varsayılsın)

Karar kuralı:

$$\text{Red } H_0 \text{ eger } \; t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} > t_{n\text{-}1,\alpha}$$

Normal dağılan bir kitlenin ortalama testi (σ Bilinmiyorsa)

(devamı)

- İki yönlü test için:

Hipotez testi aşağıdaki gibidir.

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

(kitlenin normal dağıldığı ve varyansının bilinmediği varsayılsın)

Karar kuralı:

$$\text{Red } H_0 \text{ eger } \; t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} < -t_{n\text{-}1,\alpha/2}$$

$$\text{yada eger } \; t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} > t_{n\text{-}1,\alpha/2}$$

t-tablo değeri

- http://math.stackexchange.com/questions/766122/calculate-critical-value

Örnek: İki yönlü test (σ Bilinmiyorsa)

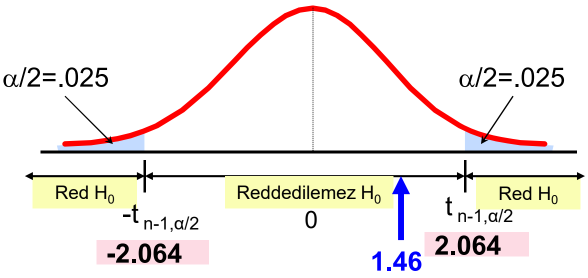
Antalya'da bulunan 5 yıldızlı bir otelde gecelik ortalama konaklama ücreti 168 TL. Rasgele 25 otel örnekleme seçiliyor ve ortalama gecelik ücreti $\bar{x}$ =172.50 TL, standart sapması s=15.40 TL olarak hesaplanıyor. α = 0.05 seviyesinde gecelik konaklama ücretinin 168 TL'den farklı olup olmadığını test ediniz.

(Kitlenin normal dağıldığını varsayınız)…

H_0: $\mu = 168$
H_1: $\mu \neq 168$

Example Solution: Two-Tail Test

H_0: μ = 168
H_1: $\mu \neq$ 168



- α = **0.05**
- **n = 25**
- **σ bilinmiyorsa, t istatistiği kullanılır**

$$t_{n-1} = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{172.50-168}{\frac{15.40}{\sqrt{25}}} = 1.46$$

- **Kritik değer:**

$t_{24,\ 0.025} = \pm\ 2.064$

H_0 reddedilemez. Konaklama ücretinin ortalama 168 TL'den farklı olduğu söylenemez.

Neler öğrendik?

- Hipotez testi kavramı
- Ortalama için z - Testi (σ biliniyorsa)
- Hipotez testinde kritik değer
- Tek kuyruk ve çift yönlü testler
- Ortalama için t - testi (σ bilinmiyorsa)

Çok Değişkenli Veri Analizi

Prof. Dr. Çağın Kandemir Çavaş

DEU Fen Fakültesi
Bilgisayar Bilimleri Bölümü

İki Örneklem Hipotez Testleri

İki Örneklem Testleri

İki Örneklem Testleri

- Kitle ortalamaları, Bağımlı örneklemler
  - Örneğin: Aynı örneklemin Önce-sonra ölçümleri karşılaştırması
- Kitle ortalamaları, Bağımsız örneklemler
  - 1.Grup ile bağımsız 2. Grup karşılaştırması

Bağımlı Örneklemler

Bağımlı Örneklemler

Normal dağılan İki Kitle ortalaması arasındaki Farkın Testi: Bağımlı Örneklemler

Bağımlı Kitlenin Ortalamalarının Testi

- Çift ya da eşleşmiş örneklemler
- Tekrarlayan ölçümler (önce/sonra)
- Çift değerler arasındaki fark hesaplanır:

$$d_i = x_i - y_i$$

- Varsayımlar:
  - Kitlelerin normal dağıldığı varsayılır.

Test İstatistiği: Bağımlı Örneklemler

Kitle ortalamaları, Bağımlı örneklemler

Aşağıdaki hipotezlerin testi için:

$H_0: \mu_x - \mu_y \geq 0$

$H_0: \mu_x - \mu_y \leq 0$

$H_0: \mu_x - \mu_y = 0$

Ortalama fark için test istatistiği $n - 1$ serbestlik dereceli t değeridir:

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

$$\bar{d} = \frac{\sum d_i}{n}$$

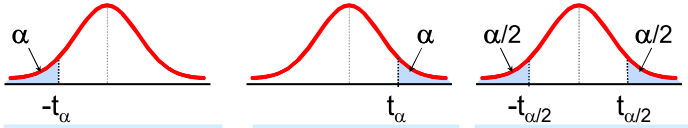
s_d = farkların örneklem standart sapması

n = örneklem büyüklüğü (eşlenmiş çift sayısı)

Karar Kuralları: Bağımlı Örneklemler

Bağımlı Örneklemler

Alt-kuyruk testi:	Üst-kuyruk testi:	İki yönlü test:
$H_0: \mu_x - \mu_y \geq 0$ $H_1: \mu_x - \mu_y < 0$	$H_0: \mu_x - \mu_y \leq 0$ $H_1: \mu_x - \mu_y > 0$	$H_0: \mu_x - \mu_y = 0$ $H_1: \mu_x - \mu_y \neq 0$



RED H_0 eğer $t < -t_{n-1,\alpha}$	RED H_0 eğer $t > t_{n-1,\alpha}$	RED H_0 eğer $t < -t_{n-1,\alpha/2}$ or $t > t_{n-1,\alpha/2}$

$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$ istatistiği n - 1 serbestlik derecesine sahiptir.

Bağımlı Örneklemler: Örnek

- Bir firma satış sorumlusu olan çalışanlarını 'Müşteri hizmetleri eğitim çalıştayına göndermiştir. Alınan eğitimle şikayet sayısında farklılık olmuş mudur? (α = 0.05) Veriler aşağıdaki gibidir:

SatışSorumlusu	Number of Complaints: Önce (1)	Sonra (2)	(2) - (1) Fark, d_i
Ali	6	4	- 2
Fırat	20	6	-14
Ayşe	3	2	- 1
Koray	0	0	0
Deniz	4	0	- 4
			-21

$$\bar{d} = \frac{\sum d_i}{n}$$

$= -4.2$

$$S_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}}$$

$= 5.67$

Bağımlı Örneklemler: Çözüm

- Alınan eğitimle şikayet sayısında farklılık olmuş mudur? (α = 0.05 anlamlılık seviyesinde)

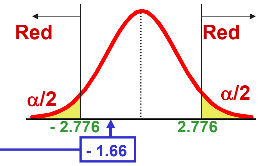
$\mathbf{H_0: \mu_x - \mu_y = 0}$

$\mathbf{H_1: \mu_x - \mu_y \neq 0}$

$\alpha = .05 \quad \bar{d} = -4.2$

Kritik değer(tablo değeri)= ± 2.776

serbestlik derecesi = n − 1 = 4



Test İstatistiği:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{-4.2}{5.67/\sqrt{5}} = \boxed{-1.66}$$

Karar: *H_0* Reddedilemez

(t test istatistiği red bölgesinde değil)

Yorum: Şikayet sayısında anlamlı bir değişiklik olmamıştır.

Bağımsız Örneklemler

Kitle ortalamaları, bağımsız örneklemler

Normal dağılan iki kitle ortalamasının arasındaki Farkın Testi: Bağımsız Örneklemler

- Farklı kitleler
  - Birbirden ayrı
  - Bağımsız
    - Bir kitleden örneklem seçilmesi, diğer kitleden örneklem seçilmesini etkilemez.
  - Normal dağılıma sahiptirler.

İki Ortalama Arasındaki Fark

(devamı)

- Kitle ortalamaları, bağımsız örneklemler
  - σ_x^2 ve σ_y^2 biliniyorsa → Test istatistiği **z** değeridir.
  - σ_x^2 ve σ_y^2 bilinmiyorsa
    - σ_x^2 ve σ_y^2 eşit varsayılırsa → Test istatistiği **t** değeridir.
    - σ_x^2 ve σ_y^2 eşit varsayılmazsa → Test istatistiği **t** değeridir.

σ_x^2 ve σ_y^2 Biliniyorsa

- Kitle ortalamaları, bağımsız örneklemler
  - σ_x^2 ve σ_y^2 biliniyorsa *
  - σ_x^2 ve σ_y^2 bilinmiyorsa

Varsayımlar:

- Örneklemler rasgele ve bağımsız seçilmiştir.
- İki kitle dağılımı da normaldir.
- Kitle varyansları biliniyor.

σ_x^2 ve σ_y^2 Biliniyorsa

(devamı)

- Kitle ortalamaları, Bağımsız örneklemler
  - σ_x^2 ve σ_y^2 biliniyorsa *
  - σ_x^2 ve σ_y^2 bilinmiyorsa

σ_x^2 ve σ_y^2 biliniyorsa ve iki kitle de normal dağılım gösteriyorsa, $\bar{X}-\bar{Y}$ farkının varyansı:

$$\sigma^2_{\bar{X}-\bar{Y}}=\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}$$

Test istatistiği:

$$Z=\frac{(\bar{x}-\bar{y})-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_x^2}{n_X}+\frac{\sigma_y^2}{n_Y}}}$$

Standart normal dağılıma sahiptir.

Test İstatistiği, σ_x^2 ve σ_y^2 Biliniyorsa

Kitle ortalamaları, Bağımsız örneklemler

- σ_x^2 ve σ_y^2 biliniyorsa *
- σ_x^2 ve σ_y^2 bilinmiyorsa

$$H_0 : \mu_x - \mu_y = 0$$

$\mu_x - \mu_y$ için test istatistiği:

$$z = \frac{(\bar{x} - \bar{y})}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}}$$

İki Kitle Ortalaması için Hipotez Testi

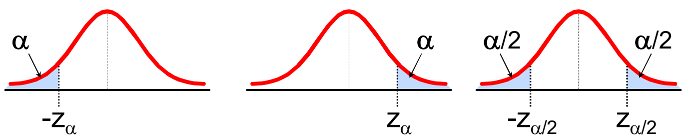
İki kitle ortalamaları, Bağımsız örneklemler

Alt-kuyruk testi:	Üst-kuyruk testi:	İki yönlü test:
$H_0: \mu_x \geq \mu_y$ $H_1: \mu_x < \mu_y$	$H_0: \mu_x \leq \mu_y$ $H_1: \mu_x > \mu_y$	$H_0: \mu_x = \mu_y$ $H_1: \mu_x \neq \mu_y$
veya,	veya,	veya,
$H_0: \mu_x - \mu_y \geq 0$ $H_1: \mu_x - \mu_y < 0$	$H_0: \mu_x - \mu_y \leq 0$ $H_1: \mu_x - \mu_y > 0$	$H_0: \mu_x - \mu_y = 0$ $H_1: \mu_x - \mu_y \neq 0$

Karar Kuralları

İki Kitle Ortalamaları, Bağımsız Örneklemler, Varyanslar biliniyorsa

Alt-kuyruk testi:	Üst-kuyruk testi:	İki yönlü test:
$H_0: \mu_x - \mu_y \geq 0$ $H_1: \mu_x - \mu_y < 0$	$H_0: \mu_x - \mu_y \leq 0$ $H_1: \mu_x - \mu_y > 0$	$H_0: \mu_x - \mu_y = 0$ $H_1: \mu_x - \mu_y \neq 0$



Red H_0 eğer $z < -z_\alpha$	Red H_0 eğer $z > z_\alpha$	Red H_0 eğer $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

Örnek

- Ankara'daki iki çocuklu ailelerin aylık mutfak harcamalarının İstanbul'daki iki çocuklu ailelerin aylık mutfak harcamalarından fazla olduğu iddia edilmektedir. Ankara ve İstanbul'da iki çocuklu ailelerden 6'şar tanesi rassal olarak seçilerek aylık mutfak harcamaları saptanmıştır.

$$X_{i1}(Ank): 30, 40, 45, 70, 60, 45$$

$$X_{i2}(\text{İst}): 35, 30, 40, 60, 40, 45$$

- Kitlelere ilişkin varyanslar $\sigma_1^2 = 25$ ve $\sigma_2^2 = 16$ olduğuna göre, $\alpha=0.05$ anlamlılık düzeyinde iddiayı test ediniz.

Çözüm

X_{i1}	X_{i2}
30	35
40	30
45	40
70	60
60	40
45	45
290	250

$$\bar{X}_1 = \frac{1}{6}(290) = 48.33$$

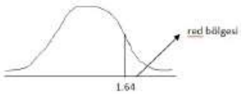
$$\bar{X}_2 = \frac{1}{6}(250) = 41.66$$

$$\bar{X}_1 - \bar{X}_2 = 6.67$$

$$\sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{25}{6} + \frac{16}{6} = 6.83$$

$$Z_h = \frac{\bar{X}_1 - \bar{X}_1}{\sqrt{\sigma^2_{\bar{X}_1-\bar{X}_2}}} = \frac{6.67}{\sqrt{6.83}} = 2.55$$

$$Z_{0.05} = 1.64$$



$Z_h > Z_{0.05}$ olduğu için H_0 reddedilir.

σ_x^2 ve σ_y^2 Bilinmiyorsa, Eşit varsayılırsa

Kitle ortalamaları, bağımsız örneklemler

- σ_x^2 ve σ_y^2 biliniyorsa
- σ_x^2 ve σ_y^2 bilinmiyorsa
  - σ_x^2 ve σ_y^2 eşit varsayılırsa *
  - σ_x^2 ve σ_y^2 eşit varsayılmazsa

Varsayımlar:

- Örneklemler rasgele ve

bağımsız seçilmiştir.

- Kitleler normal dağılıma sahiptir.
- Kitle varyansları bilinmiyor ama eşit varsayılıyor.

σ_x^2 ve σ_y^2 Bilinmiyorsa, Eşit varsayılırsa

(devamı)

Kitle ortalamaları, bağımsız örneklemler

- σ_x^2 ve σ_y^2 biliniyorsa
- σ_x^2 ve σ_y^2 bilinmiyorsa
  - σ_x^2 ve σ_y^2 eşit varsayılırsa *
  - σ_x^2 ve σ_y^2 eşit varsayılmazsa

- Kitle varyansları eşit varsayılır, örneklemlere ait iki standart sapma kullanılarak ortak bir σ hesaplanır.
- $(n_x + n_y - 2)$ serbestlik dereceli t değeri kullanılır.

Test İstatistiği, σ_x^2 ve σ_y^2 Bilinmiyorsa, Eşit varsayılırsa

σ_x^2 ve σ_y^2 bilinmiyorsa

- σ_x^2 ve σ_y^2 eşit varsayılırsa *
- σ_x^2 ve σ_y^2 eşit varsayılmazsa

Test İstatistiği:

$H_0 : \mu_x - \mu_y = 0$

$$t = \frac{(\bar{x} - \bar{y})}{\sqrt{\dfrac{s_p^2}{n_x} + \dfrac{s_p^2}{n_y}}}$$

t nin serbestlik derecesi $(n_1 + n_2 - 2)$

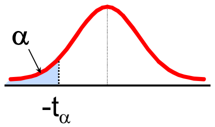
Ve ortak varyans:

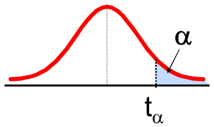
$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

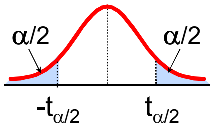
Karar Kuralları

İki kitle ortalamaları, Bağımsız Örneklemler, Varyanslar Bilinmiyorsa

Alt-kuyruk testi:	Üst-kuyruk testi:	İki yönlü test:
H_0: $\mu_x - \mu_y \geq 0$ H_1: $\mu_x - \mu_y < 0$	H_0: $\mu_x - \mu_y \leq 0$ H_1: $\mu_x - \mu_y > 0$	H_0: $\mu_x - \mu_y = 0$ H_1: $\mu_x - \mu_y \neq 0$







Red H_0 eğer
$t < -t_{(n1+n2-2),\ \alpha}$

Red H_0 eğer
$t > t_{(n1+n2-2),\ \alpha}$

Red H_0 eğer
$t < -t_{(n1+n2-2),\ \alpha/2}$ or
$t > t_{(n1+n2-2),\ \alpha/2}$

Ortak varyans Varyans t Test: Örnek

Bir borsa analiz uzmanı NYSE ve NASDAQ şirketlerine ait borsa gelirlerini incelemektedir. Aşağıdaki gibi bilgileri elde etmiştir.

	NYSE	NASDAQ
Sayısı	21	25
Örneklem Ort.	3.27	2.53
Örneklem stdsap.	1.30	1.16

İki şirketin gelirlerine ait varyansları eşit ve normal dağılıma uyduğu varsayımı altında, iki firmanın ortalama geliri arasında farklılık var mıdır? ($\alpha = 0.05$)

Test İstatistiğini Hesaplama

H_0: $\mu_1 - \mu_2 = 0$ i.e. ($\mu_1 = \mu_2$)

H_1: $\mu_1 - \mu_2 \neq 0$ i.e. ($\mu_1 \neq \mu_2$)

Test istatistiği:

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(3.27 - 2.53)}{\sqrt{1.5021\left(\frac{1}{21} + \frac{1}{25}\right)}} = \boxed{2.040}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(21 - 1)1.30^2 + (25 - 1)1.16^2}{(21 - 1) + (25 - 1)} = 1.5021$$

Çözüm

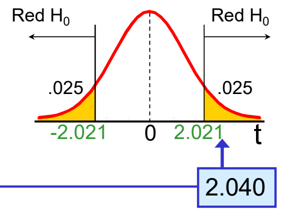
H_0: $\mu_1 - \mu_2 = 0$ veya ($\mu_1 = \mu_2$)

H_1: $\mu_1 - \mu_2 \neq 0$ veya ($\mu_1 \neq \mu_2$)

α = 0.05

Serb.der. = 21 + 25 − 2 = 44

Kritik değerler: t = ± 2.021 (yaklaşık)



Test istatistiği:

$$t = \frac{3.27 - 2.53}{\sqrt{1.5021\left(\frac{1}{21} + \frac{1}{25}\right)}} = \boxed{2.040}$$

Karar:

α = 0.05'da Red H_0

Yorum:

Ortalamalarda farklılık olduğuna dair yeterli kanıt vardır.

σ_x^2 ve σ_y^2 Bilinmiyorsa, Eşit varsayılmıyorsa

Kitle ortalamaları, bağımsız örneklemler

- σ_x^2 ve σ_y^2 biliniyorsa
- σ_x^2 ve σ_y^2 bilinmiyorsa
  - σ_x^2 ve σ_y^2 eşit varsayılırsa
  - σ_x^2 ve σ_y^2 eşit varsayılmazsa *

Varsayımlar:

- Örneklemler rasgele ve bağımsız seçilmiştir.
- Kitleler normal dağılıma uymaktadır.
- Kitle varyansları bilinmiyor ve eşit değiller.

σ_x^2 ve σ_y^2 Bilinmiyorsa, Eşit varsayılmıyorsa

(devamı)

Kitle ortalamaları, bağımsız örneklemler

- σ_x^2 ve σ_y^2 biliniyorsa
- σ_x^2 ve σ_y^2 bilinmiyorsa
  - σ_x^2 ve σ_y^2 eşit varsayılırsa
  - σ_x^2 ve σ_y^2 eşit varsayılmazsa *

- Kitle varyansları eşit kabul edilmez.
- ν serbestlik dereceli t değeri kullanılır.

$$\nu = \frac{\left[\left(\frac{s_x^2}{n_x}\right)+\left(\frac{s_y^2}{n_y}\right)\right]^2}{\left(\frac{s_x^2}{n_x}\right)^2/(n_x-1)+\left(\frac{s_y^2}{n_y}\right)^2/(n_y-1)}$$

Test İstatistiği, σ_x^2 ve σ_y^2 Bilinmiyorsa, Eşit değilse

- σ_x^2 ve σ_y^2 bilinmiyorsa
  - σ_x^2 ve σ_y^2 eşit varsayılırsa
  - σ_x^2 ve σ_y^2 eşit varsayılmazsa *

Test istatistiği:

H_0: $\mu_x - \mu_y = 0$

$$t = \frac{(\bar{x}-\bar{y})}{\sqrt{\frac{s_x^2}{n_X}+\frac{s_y^2}{n_Y}}}$$

Yeni serbestlik derecesi ν hesaplanır:

$$\nu = \frac{\left[\left(\frac{s_x^2}{n_x}\right)+\left(\frac{s_y^2}{n_y}\right)\right]^2}{\left(\frac{s_x^2}{n_x}\right)^2/(n_x-1)+\left(\frac{s_y^2}{n_y}\right)^2/(n_y-1)}$$

Örnek

Et fabrikası için çalışan bir finansal analistsiniz. İki ayrı kesimhanenin üretim kayıtlarıyla ilgili aşağıdaki verileri topladınız:

	fab1	fab2
Veri sayısı	**21**	**25**
Ortalama	**3.27**	**2.53**
Std Sapma	**1.30**	**1.16**

Kitle varyanslarının birbirine eşit olmadığı ve değerlerinin bilinmediği varsayımı altında iki fabrikanın ortalama üretiminde farklılık var mıdır? (α **= 0.05**)

MİNİTAB Uygulaması-Örnek

	Komisyoncular			
Evler	**A**	**B**		
1	181.0	182.0		
2	179.9	180.0		
3	163.0	161.5		
4	218.0	215.0		
5	213.0	216.5		
6	175.0	175.0		
7	217.9	219.5		
8	151.0	150.0		
9	164.9	165.5		
10	192.5	195.0		
11	225.0	222.7		
12	177.5	178.0		

İki komisyoncunun aynı evlere farklı fiyatlar verdiği iddia edilmektedir. İddiayı test etmek için 12 ev seçiliyor ve komisyonculardan bu evlere fiyat vermeleri isteniyor. Elde edilen sonuçlar tablodaki gibidir. İki komisyoncunun aynı evlere farklı fiyatlar verip vermediğini test ediniz.(α=0.05)

Çözüm

1.Adım: H_0: $\mu_D = 0$

H_1: $\mu_D \neq 0$

2.Adım: $\bar{D} = \frac{\sum D}{n} = \frac{-2}{12} = -0.167$ $\quad s_D = \sqrt{\frac{\sum D^2 - \frac{(\sum D)^2}{n}}{n-1}} = \sqrt{\frac{40.22 - \frac{(-2)^2}{12}}{12-1}} = 1.904$

$$t_{hes} = \frac{\bar{D}}{s_D / \sqrt{n}} = \frac{-0.167}{1.904 / \sqrt{12}} = -0.30 \qquad v = n-1 = 12-1 = 11 s.d.$$

3.Adım : t_{tab} : $t_{11,0.05}$ = ± 2.201

4.Adım : $|t_{hes}| < |t_{tab}|$

Bağımlı örneklem testidir. H_0 reddedilemez. 0.05 anlamlılık düzeyinde fiyatlandırma yönünden komisyoncuların birbirinden farklı olmadığına karar verebiliriz.

Neler öğrendik?

- İki bağımlı örneklemin testi (Eşlenmiş çift örneklemler)
  - Ortalama farkı için çift örneklem t testi
- İki bağımsız örneklemin testi
  - İki ortalamanın farkları için z testi
  - İki ortalamanın farkı için ortak varyanslı t-testi
  - İki ortalamanın farkı için yeni serbestlik dereceli t-testi

Çok Değişkenli Veri Analizi

DEÜ Fen Fakültesi
Bilgisayar Bilimleri Bölümü
Prof.Dr. Çağın Kandemir Çavaş

Kovaryans : Örnek-1

- Haftalık ders çalışma süresi (H) ile dersin sınavından alınan değerler (M) verilmiştir.

	Hours(H)	*Mark(M)*
Data	9	39
	15	56
	25	93
	14	61
	10	50
	18	75
	0	32
	16	85
	5	42
	19	70
	16	66
	20	80

- İki değişken arasındaki kovaryans nedir?

	Hours(H)	*Mark(M)*
Data	9	39
	15	56
	25	93
	14	61
	10	50
	18	75
	0	32
	16	85
	5	42
	19	70
	16	66
	20	80
Totals	167	749
Averages	13.92	62.42

H	M	$(H_i - \bar{H})$	$(M_i - \bar{M})$	$(H_i - \bar{H})(M_i - \bar{M})$
9	39	-4.92	-23.42	115.23
15	56	1.08	-6.42	-6.93
25	93	11.08	30.58	338.83
14	61	0.08	-1.42	-0.11
10	50	-3.92	-12.42	48.69
18	75	4.08	12.58	51.33
0	32	-13.92	-30.42	423.45
16	85	2.08	22.58	46.97
5	42	-8.92	-20.42	182.15
19	70	5.08	7.58	38.51
16	66	2.08	3.58	7.45
20	80	6.08	17.58	106.89
Total				1149.89
Average				104.54

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

- Değişken sayısı arttığında, hesaplanacak kovaryans sayısı

$$\frac{n!}{(n-2)!*2}$$

Örnek-2

Örnek No	x1	x2	x3
1	4	7	8
2	3	6	9
3	5	9	10
4	3	8	7
5	4	8	9

3 değişkene ait değerler verildiğine göre kovaryans matrisini bulunuz.

Korelasyon

- İki değişken arasındaki doğrusal ilişkinin ölçütüdür ve "r" harfi ile gösterilir.
- Korelasyon katsayısı -1 ile +1 arasında değişir. -1 **tam negatif ilişki**yi +1 ise **tam pozitif ilişki**yi yansıtır. o'a yaklaştıkça ilişki kuvveti azalır.

Korelasyon katsayısının hesaplanması

- Korelasyon katsayısı aşağıda verilen formülle

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} = \frac{Kov(x,y)}{\sqrt{Var(x)Var(y)}}$$

Örnek: Korelasyon matrisi

Varyans-kovaryans matrisi $S = \begin{bmatrix} 2.8 & 2.6 & 2.6 \\ 2.6 & 5.2 & 1.2 \\ 2.6 & 1.2 & 5.2 \end{bmatrix}$ ise

$$R = \begin{bmatrix} \frac{2.8}{\sqrt{(2.8)(2.8)}} & \frac{2.6}{\sqrt{(2.8)(5.2)}} & \frac{2.6}{\sqrt{(2.8)(5.2)}} \\ & \frac{5.2}{\sqrt{(5.2)(5.2)}} & \frac{1.2}{\sqrt{(5.2)(5.2)}} \\ & & \frac{5.2}{\sqrt{(5.2)(5.2)}} \end{bmatrix} \qquad R = \begin{bmatrix} 1 & 0.68 & 0.68 \\ 0.68 & 1 & 0.23 \\ 0.68 & 0.23 & 1 \end{bmatrix}$$

Örnek-3

- Ortalama vektörü,
- standart sapma,
- varyans-kovaryans matrisi
- Korelasyon matrisi

hesaplayınız.

GİSS: vücut indeksi
SKB: Kan basıncı

Kişi no	GİSS (x)	SKB(y)
1	4	12
2	11	14
3	8	11
4	15	15
5	5	11
6	16	14
7	20	15
8	9	13
9	2	10

Varyans-Kovaryans (S) ve Korelasyon (R) **Matrislerinin** Elde Edilmesi

- S ve R matrisleri matris işlemleri ile de hesaplanabilir.
- $X=[x_{11}\ x_{12}\ \dots\ x_{1p}]$ ise

$$S = \left(X'X - \bar{x}\bar{x}'n\right)/\left(n-1\right)$$

$X'X$: Kareler ve çarpımlar toplamı matrisi (KÇT)
n : örneklem sayısı

Çok Değişkenli Bağımsız İki Kitleye ilişkin Hipotez Testi (Hotelling T^2)

- **Bağımsız iki gruba** ilişkin ortalama vektörlerinin karşılaştırılması testidir.
  - Hasta ve sağlam gruplarında hemoglobin yönünden farklılık araştırılabilir.
    - Problem çok değişkenli boyuta taşınarak hemoglobin, potasyum, kalsiyum vb. değişkenler de incelenebilir.

20.03.2025 11

Hotelling T^2 (Bağımsız iki grup)

$$H_o : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix}$$

$$H_1 : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \neq \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix}$$

20.03.2025 12

Hotelling T^2 (Bağımsız iki grup)

- Hipotez testinde ortalamalar yerine **ortalama vektörleri** karşılaştırılacaktır.
- Bağımsız iki gruba ilişkin ortalama vektörlerinin karşılaştırılmasında **Hotelling T^2** test istatistiği kullanılacaktır.

Hotelling T^2 (Bağımsız iki grup)

VARSAYIMLARI

- Çok değişkenli iki ortalama vektörünün karşılaştırılmasında, her iki gruptaki verilerin ayrı ayrı normal dağılım göstermesi ;
- İki gruba ilişkin kovaryans matrislerinin homojen olması varsayımları vardır.

Hotelling T^2 (Bağımsız iki grup)

- Hipotezler aşağıdaki şekilde kurulur;

$$H_o: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix}$$

$$H_1: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \neq \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix}$$

Hotelling T^2 (Bağımsız iki grup)

$$S = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2}$$

$$T^2 = \frac{n_1 * n_2}{n_1+n_2}(\bar{x}_1-\bar{x}_2)' S^{-1}(\bar{x}_1-\bar{x}_2)$$

$$\bar{x}_1 - \bar{x}_2 = \begin{bmatrix} \bar{x}_{11} - \bar{x}_{12} \\ \vdots \\ \bar{x}_{1p} - \bar{x}_{1p} \end{bmatrix}$$ Olursa,

n_1 : 1. gruba ait gözlem sayısı
n_2 : 2. gruba ait gözlem sayısı
$\bar{x}_1$: 1. grubun ortalama vektörü
$\bar{x}_2$: 2. grubun ortalama vektörü
S_1 : 1. grubun varyans-kovaryans matrisi
S_2 : 2. grubun varyans-kovaryans matrisi
S : Ortak varyans-kovaryans matrisi

Test istatistiği;

$$F = \frac{T^2(n_1+n_2-p-1)}{(n_1+n_2-2)p}$$

$$F > F_{(p,n_1+n_2-p-1);0.05} \Rightarrow H_0 \ RED$$

Hotelling T^2

- **F tablosu**
- http://www.socr.ucla.edu/Applets.dir/F_Table.html
- $F_{3,22,0.05}=3.05$

Hotelling T^2 (Bağımsız iki grup)

- Hesaplanan F değeri F tablo değerinden büyükse "bağımsız iki ortalama vektörü arasında **fark vardır**" yorumu yapılır.
- F değeri **p ve n1+n2-p-1** serbestlik derecesine sahiptir.

Hotelling T^2 (Bağımsız iki grup)

- **Örnek-1:** Hem koroner arter hastalığı hem de diabet hastalığı tanısı konan 26 hasta ile sadece diabet tanısı konan 24 hastaya ilişkin kolestrol, şeker ve HDL değerleri verilmiştir. 3 değişken açısından hasta grupları arasında istatistiksel olarak fark var mıdır? ($\alpha=0.05$)

Örnek-1

İki hasta grubunun kolestrol, şeker ve HDL düzeyleri aşağıdaki gibidir.

1.GRUP	2.GRUP
$n_1= 26$	$n_2= 24$
ORT_Kol=262 ORT_Şek=167 ORT_HDL=37	ORT_Kol=238 ORT_Şek=169 ORT_HDL=63
Var_Kol=633 Var_Şek=577 Var_HDL=58	Var_Kol=329 Var_Şek=387 Var_HDL=30

Grupların ortak varyans-kovaryans matrisi aşağıdaki gibi verilmiştir.

$$S=\begin{bmatrix}486.7 & 159.8 & 3.9\\ 159.8 & 485.2 & -42.0\\ 3.9 & -42.0 & 45.0\end{bmatrix} \quad S^{-1}=\begin{bmatrix}0.002345 & -0.00086 & -0.00101\\ -0.00086 & 0.002558 & 0.002465\\ -0.00101 & 0.002465 & 0.024611\end{bmatrix}$$

Çözüm-1

Hipotez testini kuralım:

$$H_o : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix}$$

$$H_1 : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{bmatrix} \neq \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix}$$

Çözüm-1

Grupların ortalama vektörleri:

$$\bar{x}_1 = \begin{bmatrix} 262 \\ 167 \\ 37 \end{bmatrix} \quad \bar{x}_2 = \begin{bmatrix} 238 \\ 169 \\ 63 \end{bmatrix}$$

Ortak Varyans-Kovaryans matrisi ve tersi:

$$S = \begin{bmatrix} 486.7 & 159.8 & 3.9 \\ 159.8 & 485.2 & -42.0 \\ 3.9 & -42.0 & 45.0 \end{bmatrix} \qquad S^{-1} = \begin{bmatrix} 0.002345 & -0.00086 & -0.00101 \\ -0.00086 & 0.002558 & 0.002465 \\ -0.00101 & 0.002465 & 0.024611 \end{bmatrix}$$

$$T^2 = \frac{26*24}{26+24} * \begin{bmatrix} 262-238 & 167-169 & 37-63 \end{bmatrix} * S^{-1} * \begin{bmatrix} 262-238 \\ 167-169 \\ 37-63 \end{bmatrix}$$

$$T^2 = 240.63$$

$$\boxed{F = \frac{240.63(26+24-3-1)}{(26+24-2)3} = 76.868} \longrightarrow \text{Test istatistiği}$$

Çözüm-1

$p = 3$

$n_1 = 24$

$n_2 = 26$

$F_{tablo(3,46,\alpha=0.05)} \cong 2.84$

$$F > F_{(3,46);0.05} \Rightarrow \quad H_0 \ RED$$

YORUM: 3 değişken açısından iki hasta grupları arasında istatistiksel olarak birbirinden farklılık bulunuyor.

Örnek-2

- Asya dağlık A ve B bölgesinde yaşayan 2 yaş çocuklarının boy(cm), göğüs kafesi çevresi(cm) ve kol pazı çevresi (cm) olarak ölçülmüştür. İki bölgenin ölçümlerinin istatistiksel olarak farklı olup olmadığını test ediniz.(α=0.05)

A Bölgesi

Birey No	Boy (cm)	Göğüs Çevresi (cm)	Kol pazu çevresi (cm)
1	78	60.6	16.5
2	76	58.1	12.5
3	92	63.2	14.5
4	81	59.0	14.0
5	81	60.8	15.5
6	84	59.5	14.0

B Bölgesi

Birey No	Boy (cm)	Göğüs Çevresi (cm)	Kol pazu çevresi (cm)
1	80	58,4	14
2	75	59,2	15
3	78	60,3	15
4	75	57,4	13
5	79	59,5	14
6	78	58,1	14,5
7	75	58	12,5
8	64	55,5	11
9	80	59,2	12,5

Çözüm-2

$$H_0: \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix}$$

$$H_1: \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{bmatrix} \neq \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix}$$

- Veri matrislerinin ortalama vektörü ve kovaryans matrisleri hesaplanır.
- Veri matrisinin ortak varyans-kovaryans matrisi bulunur (S)
- Varyans-kovaryans matrisinin tersi bulunur. (S^{-1})
- T2 test istatistiği hesaplanır.
- T2'nin F dönüşümü elde edilir.
- Hesaplanan F değeri, (p,n1+n2-p-1) serbestlik dereceli F tablo değeri ile karşılaştırılır F>F_tablo ise RED

Çözüm-2

$$\bar{x}_1 = [82.0 \quad 60.2 \quad 14.5]$$
$$\bar{x}_2 = [76.0 \quad 58.4 \quad 13.5]$$

$$S_1 = \begin{bmatrix} 31.6 & 8.040 & 0.500 \\ 8.040 & 3.172 & 1.310 \\ 0.500 & 1.310 & 1.900 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 24.5 & 5.64 & 4.31 \\ 5.64 & 1.97 & 1.45 \\ 4.31 & 1.45 & 1.81 \end{bmatrix}$$

$$S = \begin{bmatrix} 27.23 & 6.56 & 2.84 \\ 6.56 & 2.43 & 1.40 \\ 2.84 & 1.40 & 1.84 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} 0.115 & -0.37 & 0.103 \\ -0.37 & 1.92 & -0.88 \\ 0.103 & -0.88 & 1.05 \end{bmatrix}$$

Çözüm-2

$$T^2 = \frac{n_1 * n_2}{n_1 + n_2}(\bar{x}_1 - \bar{x}_2)' S^{-1}(\bar{x}_1 - \bar{x}_2)$$

$$F* = \frac{(n_1 + n_2 - p - 1)}{(n_1 + n_2 - 2)p} T^2 \sim F(p, n_1 + n_2 - p - 1)$$

$$F* > F_{(p, n_1 + n_2 - p - 1);\alpha}$$

Çözüm-2

$$T^2 = 5.31$$
$$F = 1.498$$
$$F_{(3,11);0.05} = 3.58$$

$$F < F_{(3,11);0.05} \Rightarrow \quad H_0 \ REDDEDILEMEZ$$

YORUM: Asya dağlık A ve B bölgesinde yaşayan 2 yaşındaki çocukların boy, göğüs çevresi ve kol-pazu çevresi değerleri istatistiksel olarak farklı değildir.

ÇOK DEĞİŞKENLİ VERİ ANALİZİ

Prof.Dr. Çağın Kandemir Çavaş
DEÜ Fen Fakültesi
Bilgisayar Bilimleri Bölümü

Çok Değişkenli **Bağımlı** İki Kitleye ilişkin Hipotez Testi (Hotelling T²)

$H_0: \mu_D = 0$

$H_A: \mu_D \neq 0$ veya

$H_A: \mu_D < 0$

veya

$H_A: \mu_D > 0$

$$T^2 = n(\bar{d} - \mu_0)' S_{\bar{d}}^{-1} (\bar{d} - \mu_0)$$

$$F = \frac{n-p}{(n-1)p} T^2$$

$F_{tablo} = F_{p;(n-p);\alpha}$

KARAR:

$F > F_{tablo}$ → Ho red

$\mu_0 = hipotezdeki\ eşitliğin\ sağ\ tarafı$

$\bar{x}_d$ veya $\bar{d}$: $ortalama\ fark\ vektörü$

$S_{\bar{d}}$: $farkların\ varyans - kovaryans\ matrisi$

$S_{\bar{d}}^{-1}$: $farkların\ varyans - kovaryans\ matrisinin\ tersi$

p: $değişken\ sayısı$

n: $gözlem\ sayısı$

Örnek

- 10 hastanın A ilacı ile tedavi öncesi 4 değişkenine ilişkin (x1,x2,x3 ve x4) ölçümleri ve A ilacı ile tedavi sonrası aynı değişkenlere ilişkin değerlerin ölçümler aşağıda verilmiştir. A ilacının dört değişken üzerine etkilerinin önemli olup olmadığını test ediniz. (α=0.05)

A(-)

x1	x2	x3	x4
15	8	7	11
13	7	6	12
16	9	9	13
19	8	10	14
15	9	12	14
12	10	14	20
17	11	16	22
14	8	11	16
16	9	13	15
18	11	14	16

A(+)

x1	x2	x3	x4
16	9	9	16
12	11	10	13
15	18	14	15
16	17	17	25
11	10	13	19
15	9	17	20
16	10	12	18
17	11	16	19
16	13	12	21
20	16	15	27

Çözüm

$$H_0 : \mu_d = 0 \longrightarrow \mu_0$$
$$H_1 : \mu_d \neq 0$$

FARKLAR MATRİSİ

$$M1 = \begin{bmatrix} 15 & 8 & 7 & 11 \\ 13 & 7 & 6 & 12 \\ 16 & 9 & 9 & 13 \\ 19 & 8 & 10 & 14 \\ 15 & 9 & 12 & 14 \\ 12 & 10 & 14 & 20 \\ 17 & 11 & 16 & 22 \\ 14 & 8 & 11 & 16 \\ 16 & 9 & 13 & 15 \\ 18 & 11 & 14 & 16 \end{bmatrix} \quad M2 = \begin{bmatrix} 16 & 9 & 9 & 16 \\ 12 & 11 & 10 & 13 \\ 15 & 18 & 14 & 15 \\ 16 & 17 & 17 & 25 \\ 11 & 10 & 13 & 19 \\ 15 & 9 & 17 & 20 \\ 16 & 10 & 12 & 18 \\ 17 & 11 & 16 & 19 \\ 16 & 13 & 12 & 21 \\ 20 & 16 & 15 & 27 \end{bmatrix} \quad \mu_d = \begin{bmatrix} -1 & -1 & -2 & -5 \\ 1 & -4 & -4 & -1 \\ 1 & -9 & -5 & -2 \\ 3 & -9 & -7 & -11 \\ 4 & -1 & -1 & -5 \\ -3 & 1 & -3 & 0 \\ 1 & 1 & 4 & 4 \\ -3 & -3 & -5 & -3 \\ 0 & -4 & 1 & -6 \\ 2 & -5 & -1 & -11 \end{bmatrix}$$

Çözüm

$$S = \begin{bmatrix} 5.6556 & -2.5111 & 0.0333 & -1.3333 \\ -2.5111 & 12.9333 & 7.3111 & 9.6667 \\ 0.0333 & 7.3111 & 10.4556 & 5.5556 \\ -1.3333 & 9.6667 & 5.5556 & 22.0000 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} 0.2077 & 0.0727 & -0.0476 & -0.0073 \\ 0.0727 & 0.1903 & -0.1054 & -0.0526 \\ -0.0476 & -0.1054 & 0.1691 & 0.0007 \\ -0.0073 & -0.0526 & 0.0007 & 0.0679 \end{bmatrix}$$

$$\bar{x}_d = [0.1000 \quad -3.4000 \quad -2.3000 \quad -4.0000]$$

$$T^2 = 10 * (\bar{x}_d - \mu_d) * S^{-1} * (\bar{x}_d - \mu_d)' = 10.9634$$

$$F = \frac{T^2 (n-p)}{(n-1)p} = 1.8272$$

H_0 Reddedilemez

Tek Yönlü Varyans Analizi (ANOVA)

- Tek yönlü varyans analizi, iki ya da daha fazla bağımsız ve bağımlı gruplarda çok değişkenli normal dağılımlara dayalı hipotezleri test eder.
- "ANalysis Of VAriance" ın kısaltmasıdır.
- Grup sayısı 2'den fazla olduğunda Hotelling T^2 **çalışmaz!**

ANOVA

	İşlemler/Gruplar					
	1	2	…	i	…	k
	X_{11}	X_{21}	…	X_{i1}	…	X_{k1}
	X_{12}	X_{22}	…	X_{i2}	…	X_{k2}
	.					
	.					
	X_{1n}	X_{2n}	…	X_{in}	…	X_{kn}
Toplam	T_1	T_2		T_i	T_k	T
Ortalama						

K adet grubun ortalamalarını karşılaştırabiliriz.

ANOVA

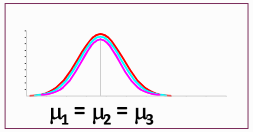
- **Hipotez Testleri:** Kurulabilecek sıfır hipotezi ve alternatif hipotez aşağıdaki gibi olur.

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

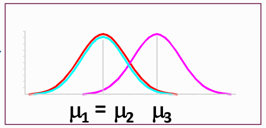
$$H_1 : En \text{ az iki anakütle ortalaması birbirine eşit değildir}$$

ANOVA

- ***H*₀**: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_c$
  - Tüm grup ortalamaları eşittir.
  - (Tedavi etkisi yoktur.)



- **H_1: Tüm μ_j ler eşit değildir.**
  - Gruplardan en az birinin ortalaması diğerlerininkinden farklıdır.
  - (Tedavi etkisi vardır.)



ANOVA

- Bu analiz sonunda, örnekler arasında uygunluk olup olmadığı yani söz konusu örneklerin aynı anakütleye ait birer rassal örnek olup olmadıkları araştırılır.

ANOVA- Genel Formül

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij}-\overline{X})^2 = \sum_{i=1}^{k} n(\overline{X}_i-\overline{X})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij}-\overline{X}_i)^2$$

Genel Kareler Toplamı **(GKT)**

Gruplar Arası Kareler Toplamı **(GAKT)**

Grup içi Kareler Toplamı **(GİKT)**

X_{ij} : i. grup j. gözlem değeri

$\overline{X}$: tüm gözlem değerlerinin ortalaması

$\overline{X}_i$: i. grubun ortalaması

k : grup sayısı

n : her bir gruptaki gözlem sayısı

ANOVA

- Aşağıda ANOVA tablosu verilmiştir.

Değişim Kaynağı	**Kareler Toplamı**	**Serbestlik Derecesi**	**Kareler Ortalaması**	**Test İstatistiği**
işlem (GrupArası)	GAKT $\sum_{i=1}^{k} n(X_i - \overline{X})^2$	**v_1=k-1**	$s_1^2 = \frac{GAKT}{k-1}$	$F = \frac{s_1^2}{s_2^2}$
Hata (Grupİçi)	GİKT $\sum_{i=1}^{k}\sum_{j=1}^{n} (X_{ij} - \overline{X}_i)^2$	**v_2= N-k**	$s_2^2 = \frac{GİKT}{N-k}$	
Toplam	GKT $\sum_{i=1}^{k}\sum_{j=1}^{n} (X_{ij} - \overline{X})^2$	N-1		

KARAR:

$F > F_{tablo}$ → Ho RED

N: toplam gözlem sayısı

$F_{tablo} = F_{k-1;\ N-k;\ \alpha}$

ANOVA-Örnek

- Kimyevi gübre türünün dekardan alınan verimi etkileyip etkilemediği ortaya konulmak isteniyor.

Nitrat	Potasyum	Fosfat	Amonyum
20	13	5	11
17	15	4	11
14	14	13	9
19	12	18	8
16	12	10	7
16	6	10	8
102	72	60	54

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$_1 : \mu_i \neq \mu_j \quad (i \neq j)$$

Değişim Kaynağı	**Kareler Toplamı**	**Serbestlik Derecesi**	**Kareler Ortalaması**	**Test İstatistiği**
Gruplar arası	228	**3**	76	F=76/11.1 =6.8468
Hata (Gruplar içi)	222	**20**	11.1	
Toplam	450	23		

$$F > F_{(3,20);0.05} \quad \Rightarrow \quad H_0 \ RED$$

ANOVA

- Hesaplanan F değeri, F tablosundan elde edilen kritik değerden küçükse örnek ortalamaları arasındaki farklılık rassaldır; yani şanstan ileri gelmiştir ve örnekler aynı kitleden toplanmıştır.

Çok Değişkenli Veri Analizi

Prof. Dr. Çağın Kandemir Çavaş

DEU Fen Fakültesi
Bilgisayar Bilimleri Bölümü

ANOVA-Örnek

Üç sınıfa ilişkin alınan puanlar yanda verilmektedir. Buna göre sınıf ortalamalarının birbirinden farklılık gösterip göstermediğini test ediniz. ($\alpha = 0.05$)

A-1	A-2	A-3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

ANOVA-Örnek

A-1	A-2	A-3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

$\bar{x}_1 = 249.2$ | $\bar{x}_2 = 226.0$ | $\bar{x}_3 = 205.8$

$\bar{\bar{x}} = 227.0$

ANOVA-Örnek

A-1	A-2	A-3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

$\bar{x}_1 = 249.2 \quad n_1 = 5$

$\bar{x}_2 = 226.0 \quad n_2 = 5$

$\bar{x}_3 = 205.8 \quad n_3 = 5$

$\bar{\bar{x}} = 227.0$

$n = 15$

$K = 3$

GAKT = $5\,(249.2 - 227)^2 + 5\,(226 - 227)^2 + 5\,(205.8 - 227)^2 = 4716.4$

GİKT = $(254 - 249.2)^2 + (263 - 249.2)^2 + \ldots + (204 - 205.8)^2 = 1119.6$

$s_1^2 = 4716.4 / (3-1) = 2358.2$

$s_2^2 = 1119.6 / (15-3) = 93.3$

$$F = \frac{2358.2}{93.3} = 25.275$$

ANOVA-Örnek

H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: μ_i eşit değil

$\alpha = .05$

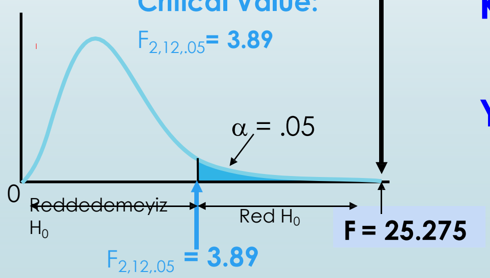
$sd_1 = 2$ $sd_2 = 12$

Test Statistic:

$$F = \frac{MSA}{MSW} = \frac{2358.2}{93.3} = 25.275$$

Critical Value:

$F_{2,12,.05} = 3.89$



Karar:

$\alpha = 0.05$ 'da H_0 Red

Yorum:

Ortalamalardan en az biri diğerlerinden farklıdır.

MANOVA

- Çok değişkenli varyans analizidir.
- '**M**ultivariate **AN**alysis **O**f **VA**riance' ın kısaltmasıdır.

6

MANOVA

- Tek yönlü varyans analizi (ANOVA) bir değişken açısından 2'den çok bağımsız grup arasında fark olup olmadığını test ediyordu.

Yüzme	Atletizm	Futbol
56	63	54
62	67	58
.	.	.
69	65	61

Yüzme, atletizm ve futbol sporları ile uğraşan kişilerin ortalama oksijen tüketim miktarları karşılaştırılıyor.

MANOVA

- Karşılaştırılmak istenen gruplara ilişkin farklı özellikler olduğunda, mesela 3 farklı ölçüm değeri karşılaştırılmak istendiğinde MANOVA kullanılır.

Yüzme			Atletizm			Futbol		
Max	MKK	SBİL	Max	MKK	SBİL	Max	MKK	SBİL
56	44	62	63	38	65	54	51	36
62	48	46	67	41	57	58	49	42
.	.	.	.	.	.	.	.	.
69	36	63	65	33	44	61	54	51

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$

H_1 : En azından bir ortalama diğerinden farklıdır.

MANOVA

VARSAYIMLAR:

- Grupların varyans-kovaryans matrislerinin homojendir.
- Her grup çok değişkenli normal dağılım gösterir.

$$H_0 : \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{01} \\ \boldsymbol{\mu}_{02} \\ \vdots \\ \boldsymbol{\mu}_{0k} \end{bmatrix} \qquad H_1 : \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_k \end{bmatrix} \neq \begin{bmatrix} \boldsymbol{\mu}_{01} \\ \boldsymbol{\mu}_{02} \\ \vdots \\ \boldsymbol{\mu}_{0k} \end{bmatrix}$$

Kullanılan test istatistiğinin mantığı tek yönlü varyans analizinden yola çıkılarak geliştirilmiştir.

9

Kullanılan Test İstatistiği ve Hipotezler

- Çok değişkenli analizlerde ortalama vektörleri arasında fark olup olmadığını incelemek için kullanılan test istatistiği vardır.

. **Wilk's Lamda İstatistiği**

10

Kullanılan Test İstatistikleri ve Hipotezler

- Çok değişkenli varyans analizinde hipotez ikiden çok ortalama vektörünün eşit olduğu şekilde kurulur.

MANOVA yorum

- Hesaplanan gruplar arası kareler toplamının grup içi kareler toplamına **oranı küçük** ise gruplar arasında **fark yoktur**.
- Bu **oran büyük** ise gruplar arasında **fark vardır**.
- *Grup içi kareler toplamının genel kareler toplamına oranı 1'e yaklaşıyorsa gruplar arası fark yok, 0'a yaklaşıyorsa fark vardır yorumu yapılır.*

GKT = GAKT + GİKT

Wilks Lambda İstatistiği

- Varyans analizinde

GKT=GAKT+GİKT

şeklinde verilen eşitlik çok değişkenli analizlerde matrislerle ifade edilir.

T (GKT): Genel Kareler Toplamı Matrisi

B (GAKT): Gruplar Arası Kareler Toplamı Matrisi

W (GİKT):Grup içi Kareler Toplamı Matrisi

p: Her gruptaki değişken sayısı

k: Grup sayısı

13

Wilks Lambda İstatistiği

$$T = B + W$$

$$T = (N-1)S$$

$$B = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'$$

$$W = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'$$

$$W = \sum_{i=1}^{k}(n_i - 1)S_i$$

k = Grup Sayısı

$\bar{x}_i$ = i. Gruba ait ortalama vektörü

$\bar{x}$ = Genel ortalama vektörü

n_i = i. Gruba ait gözlem sayısı

S_i = i. Gruba ilişkin varyans-kovaryans matrisi

S = Tüm verilerin varyans-kovaryans matrisi

$$\Lambda = \frac{|W|}{|B+W|}$$

14

Wilks Lambda İstatistiği

- Hesaplanan değer determinantların oranını içermektedir.

$$\Lambda = \frac{|W|}{|B+W|}$$

Bu oranın 0'a yaklaşması gruplar arasında fark olduğunun göstergesidir.

15

Test İstatistiği (L)

- k>2 ve p>1 Toplam gözlem sayısı (N>30) yeteri kadar büyükse:

$$L = -\left(N - 1 - \frac{p+k}{2}\right)\ln\Lambda$$

Hesaplanan değer p(k-1) serbestlik dereceli X^2 dağılımı gösterir.

$L > X^2_{tablo[p(k-1);a]}$

ise gruplar arası fark vardır denir.

- k>2 ve p>1 ve Toplam gözlem sayısı (N<30) yeteri kadar küçükse:

$$L = -\left(\frac{N-p-2}{p}\right)\left(\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\right)$$

Hesaplanan değer 2p,2(N-p-2) serbestlik dereceli F dağılımı gösterir.

$L > F_{(2p,2(N-p-2));a}$

ise gruplar arası fark vardır denir.

χ^2 (Kİ- KARE) TABLOSU

http://www.di-mgt.com.au/chisquare-table.html

MANOVA Örnek-1:

- **Üç** ayrı bölgede yaşayan çocukların **hemoglobin** (HG) ve **kan protein** (KP) değerleri tabloda verilmiştir. Bölgeler arasında bu iki değişken açısından fark var mıdır?

18

Hipotez Kurulumu:

BÖLGELER					
A		B		C	
HG	KP	HG	KP	HG	KP
12,3	5,2	12,0	5,9	11,4	4,2
11,5	4,0	11,6	5,3	11,4	5,6
11,7	4,7	11,4	5,8	11,0	5,8
11,5	5,3	11,8	5,9	11,1	4,4
11,8	5,6	12,0	5,3	12,5	4,7
12,0	5,4	12,1	4,8	12,3	4,2
12,2	4,8	11,1	4,6	10,3	5,3
10,8	4,9	11,8	4,0	11,6	5,8
11,8	4,1	11,2	4,1	12,0	5,7
12,4	5,3	12,0	5,9	11,8	5,6
12,8	5,8	12,4	4,3		
		11,6	4,6		

H_0: $\mu_A = \mu_B = \mu_C$

H_1: $\mu_A \neq \mu_B \neq \mu_C$

$$H_0: \begin{bmatrix} HG \\ KP \end{bmatrix}_A = \begin{bmatrix} HG \\ KP \end{bmatrix}_B = \begin{bmatrix} HG \\ KP \end{bmatrix}_C$$

$$H_0: \begin{bmatrix} HG \\ KP \end{bmatrix}_A \neq \begin{bmatrix} HG \\ KP \end{bmatrix}_B \neq \begin{bmatrix} HG \\ KP \end{bmatrix}_C$$

- Çözüm:

$$\bar{x}_1 = \begin{bmatrix} 11.89 \\ 5.01 \end{bmatrix} \quad \bar{x}_2 = \begin{bmatrix} 11.75 \\ 5.04 \end{bmatrix} \quad \bar{x}_3 = \begin{bmatrix} 11.54 \\ 5.13 \end{bmatrix} \quad \bar{x} = \begin{bmatrix} 11.73 \\ 5.06 \end{bmatrix}$$

$$B = 11 \begin{bmatrix} 11.89 - 11.73 \\ 5.01 - 5.06 \end{bmatrix} \begin{bmatrix} 11.89 - 11.73 & 5.01 - 5.06 \end{bmatrix}$$
$$+ \; 12 \begin{bmatrix} 11.75 - 11.73 \\ 5.04 - 5.06 \end{bmatrix} \begin{bmatrix} 11.75 - 11.73 & 5.04 - 5.06 \end{bmatrix}$$
$$+ \; 10 \begin{bmatrix} 11.54 - 11.73 \\ 5.13 - 5.06 \end{bmatrix} \begin{bmatrix} 11.54 - 11.73 & 5.13 - 5.06 \end{bmatrix}$$

$$= \begin{bmatrix} 0.650 & -0.227 \\ -0.227 & 0.081 \end{bmatrix} \longrightarrow$$ **Gruplararası Kareler Toplamı**

$$S_1 = \begin{bmatrix} 0.291 & 0.141 \\ 0.141 & 0.333 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 0.148 & 0.044 \\ 0.044 & 0.535 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 0.427 & -0.098 \\ -0.098 & 0.460 \end{bmatrix}$$

Grupların varyans-kovaryans matrisi

$$W = \sum_{i=1}^{k} (n_i - 1) S_i \qquad W = \begin{bmatrix} 8.383 & 1.014 \\ 1.014 & 13.36 \end{bmatrix} \longrightarrow$$

Grupiçi Kareler Toplamı

$$T = B + W = \begin{bmatrix} 9.033 & 0.79 \\ 0.79 & 13.44 \end{bmatrix} \longrightarrow$$

Toplam Kareler Toplamı

Lambda değeri

$$|W| = 110.986 \qquad \Lambda = \frac{110.986}{120.779} = 0.919$$

$$|B + W| = 120.779$$

$$L = -\left[33 - 1 - \frac{2+3}{2}\right] \ln(0.919)$$

$$= 2.4918 \longrightarrow$$ **Wilks Lambda İstatistiği**

$$L = 2.4918 < X^2_{Tablo\,(2(3-1);0.05)}$$

$$9.488$$

Ho reddedilemez. Hesaplanan L değeri tablo değerinden küçük olduğundan dolayı üç ortalama vektörü arasında fark yoktur sonucuna varılır.